

Development and external validation of competing risk models for major osteoporotic fracture and hip fracture in an ethnically diverse UK population of 15.6 million adults

Julia Hippisley-Cox

julia.hippisley-cox@qmul.ac.uk

Queen Mary University of London <https://orcid.org/0000-0002-2479-7283>

Carol Coupland

Queen Mary University of London

Article

Keywords:

Posted Date: September 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10828533/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations:

Yes there is potential Competing Interest. The authors declare the following competing interests: JHC reports grants from National Institute for Health Research, and other research councils, during the conduct of the study. JHC is director of QResearch, a not-for-profit unit based at Queen Mary University of London. Until 2023, JHC had a 50% shareholding in ClinRisk Ltd, co-owning it with her husband, who was an executive director. In 2023, 100% of the share capital was donated to Endeavour Health Care Charitable Trust and the company renamed to Endeavour Predict Ltd. JHC is a consultant to Endeavour Predict Ltd. and her husband is a non-executive director. The company licences software (including QFracture) both to the private sector and to NHS bodies or bodies that provide services to the NHS (through GP electronic health record providers, pharmacies, hospital providers and other NHS providers). CC reports receiving personal fees from ClinRisk Ltd until 2023, outside this work.

I can confirm that the study was approved by the East Midlands–Derby Research Ethics Committee for QResearch (OX330, reference 23/EM/0400) - this is actually mentioned in the manuscript on page 32. The study used anonymised data and hence no individual patient consent was required.

Development and external validation of competing risk models for major osteoporotic fracture and hip fracture in an ethnically diverse UK population of 15.6 million adults

Authors

Julia Hippisley-Cox¹

Carol AC Coupland²

Affiliations

1 Professor of Clinical Epidemiology & Predictive Medicine. Queen Mary University of London, UK. <http://orcid.org/0000-0002-2479-7283>

2 Senior Research Fellow Queen Mary University of London, UK. <https://orcid.org/0000-0002-2327-3306>

Corresponding author: julia.hippisley-cox@qmul.ac.uk

ABSTRACT

Fragility fractures are a major cause of mortality, disability and loss of independence in adults worldwide, yet existing fracture-risk prediction algorithms incompletely capture the growing populations of individuals living with learning disability, chronic disease, HIV and cancer survivorship.

We developed sex-specific prediction models of 10-year major osteoporotic and hip-fracture risk in 9,465,613 UK adults in primary care and evaluated them in held-out QResearch (n = 3,155,655) and external CPRD (n = 3,011,531) cohorts. The corresponding incident numbers of major osteoporotic fractures during follow-up were 166,442, 55,413 and 55,145.

Cause-specific Cox models estimated risk with competing mortality. Model A re-estimated established QFracture predictors; Model B additionally incorporated learning disability, Down syndrome, multiple sclerosis, visual impairment, HIV, selected cancers and cancer-directed endocrine therapies. Discrimination was high across validation cohorts (C-statistics, 0.766–0.900). Models A and B showed better overall calibration than QFracture and higher estimated net benefit over parts of the prespecified threshold range; residual miscalibration remained in some external-validation strata. Most population-level improvement arose from contemporary estimation within the competing risk framework. Added predictors produced clinically informative reprioritization of selected individuals.

These models could enable more accurate, inclusive fracture-risk assessment.

INTRODUCTION

Fragility fractures are a major cause of mortality, disability and loss of independence worldwide. An estimated 37 million fragility fractures occur annually among adults aged over 55 years, and their incidence and health care consequences are expected to increase as populations age.^{1,2} Fracture-risk prediction models are used to guide bone-mineral-density assessment, preventive treatment and specialist referral.³⁻⁶ Although effective preventive treatments are available, osteoporosis commonly remains undiagnosed until a first fracture occurs. Identifying individuals at high risk before that event is therefore an important clinical and public-health priority.¹

Population studies have demonstrated increased fracture risks among people with learning disability, selected cancers, visual impairment, heart failure, multiple sclerosis and HIV.⁷⁻¹³ Aromatase inhibitors and androgen-deprivation therapy may further increase skeletal risk among cancer survivors.¹¹ These characteristics are not explicitly represented in the core predictor architecture of FRAX or QFracture, two widely used fracture-prediction tools internationally.¹⁴ Improved survival has contributed to growing populations living with disability, HIV/AIDS, cancer and complex chronic disease, but it remains uncertain whether explicitly incorporating these characteristics improves fracture-risk prediction beyond established predictors. Some affected individuals may also encounter barriers to bone-health assessment. More inclusive representation of clinical risk factors in routinely implemented prediction tools has the potential to improve equitable case finding, provided that it produces clinically useful risk estimates.

Accurate estimation of absolute fracture risk also requires consideration of competing mortality, particularly among older people and those with multimorbidity^{4,15}. Failure to account adequately for the probability of death before fracture may overestimate 10-year fracture risk and alter who is selected for assessment or preventive intervention. It is therefore important to distinguish improvements arising from contemporary estimation and competing-risk modelling from any incremental value contributed by additional predictors.

We therefore developed and externally validated sex-specific models for predicting 10-year risks of major osteoporotic fracture and hip fracture in 15.6 million adults aged 30–100 years using linked UK primary-care, hospital-admission and mortality records. The models are intended to support routine primary-care decisions made by clinicians about further fracture-risk assessment, bone-density testing and preventive intervention. To distinguish the contribution of improvements to risk framework from that of the additional predictors, Model A contemporaneously re-estimated the established QFracture predictor architecture while accounting for competing mortality, whereas Model B additionally considered disability, neurological disease, HIV, selected cancers and cancer-directed endocrine therapies. In a separate ethnicity-omission analysis, we compared Model B with Model C, which excluded ethnicity, to inform current international debates concerning ethnicity and algorithmic fairness in fracture-risk assessment¹⁶⁻²⁰. We evaluated discrimination, calibration, prognostic accuracy, clinical utility and reclassification in a held-out QResearch cohort and an external CPRD cohort.

RESULTS

Study Population

The derivation cohort comprised 9,465,613 participants from England, the held-out QResearch validation cohort included 3,155,655 participants from England, and the external CPRD validation cohort included 3,011,531 participants from England and Wales (Table 1). The corresponding incident numbers of major osteoporotic fractures during follow-up were 166,442, 55,413 and 55,145, and of hip fractures were 85,174, 27,969 and 31,915 respectively. Extended Data Table 1 shows the flow of subjects and the relevant exclusions.

Baseline characteristics were broadly similar across cohorts, although ethnicity was more completely recorded in the two QResearch cohorts (>93%) than in the external CPRD validation cohort (27.2%). Mean age ranged from 51.2 to 51.6 years, approximately half of participants were women, and the prevalence of established fracture risk factors was comparable between cohorts. The study cohorts included large numbers of individuals with several newly incorporated predictors, including learning disability, Down syndrome, heart failure, multiple sclerosis, HIV/AIDS, visual impairment and specific cancers.

Rates of major osteoporotic fracture and hip fracture increased steeply with age and were consistently higher in women than in men (Extended Data Figure 1). Rates were similar in the external CPRD cohort compared with the QResearch cohorts for both major osteoporotic fracture and hip fracture although were slightly higher at the extremes of age.

Major osteoporotic fracture

Adjusted hazard ratios for major osteoporotic fracture in the derivation cohort from the cause-specific models including new predictor variables are shown in Figure 1. Established predictors were retained in the models for men and women including previous osteoporotic fracture, alcohol and smoking status, corticosteroid exposure, rheumatoid disease, falls and Parkinson disease. Heuristic shrinkage estimates were 0.99 indicating little apparent over-fitting.²²

Among women, aromatase-inhibitor treatment, learning disability, Down syndrome, heart failure, multiple sclerosis, visual impairment and four cancer types (blood, cervical, colorectal and lung) were retained as additional predictors of increased osteoporotic fracture risk. Among men, additional predictors included androgen-deprivation therapy, learning disability, Down syndrome, heart failure, multiple sclerosis, visual impairment and six cancer types (prostate, blood, pancreatic, colorectal, liver and lung). Several associations varied with age, with larger relative effects at younger ages for predictors included in age-interaction terms (Extended Data Figure 2).

Hip fracture

Figure 2 shows the corresponding results for hip fracture. Associations were stronger than for major osteoporotic fracture, particularly for previous fracture, dementia, Parkinson's disease, visual impairment, care-home residence and selected malignancies. Strong

prognostic associations were observed for Down syndrome, multiple sclerosis, and lung cancer in men and women and for HIV/AIDS and androgen-deprivation therapy in men. Several predictors showed age-dependent effects, with larger relative effects at younger ages (Extended Data Figure 3).

The corresponding results for non-fracture deaths are shown in Supplementary Figures 1 and 2.

Ethnicity

The derivation sample included 1,857,069 non-White participants (Table 1). Marked differences in fracture risk were observed between ethnic groups (Extended Data Figure 4). Compared with White individuals, risks of both fracture outcomes were consistently lower among Black African, Caribbean, Indian, Bangladeshi, Pakistani, Chinese and other Asian populations in both sexes. The lowest adjusted hazard ratios were observed among Black African, Caribbean and Bangladeshi populations.

Absolute predicted risk trajectories

To illustrate the clinical implications of including selected additional predictors, we estimated model-based 10-year absolute risks of major osteoporotic fracture and hip fracture across ages 50–85 years for otherwise comparable individuals with and without each predictor (Figure 3). Predictions were generated for a reference individual of White ethnicity with a body mass index of 25 kg/m², who was a non-smoker and non-drinker and had no additional adverse clinical predictors, with all other characteristics held constant. These trajectories are illustrative model-based predictions.

Several additional predictors shifted predicted risk trajectories towards illustrative 3% and 10% thresholds at younger ages than in reference individuals. For example, women and men with learning disability reached the illustrative 10% threshold for major osteoporotic fracture approximately 10 years earlier than otherwise comparable individuals without learning disability. Men prescribed androgen-deprivation therapy and men with Down syndrome reached the illustrative 3% threshold for hip fracture approximately 10 and 20 years earlier, respectively. The trajectories also illustrate the importance of accounting for competing risks and modelling age-dependent effects, because the relative contribution of several predictors varied substantially across the age range. Patterns were similar for other scenarios.

Discrimination, calibration and decision curve analysis

Table 2 compares discrimination and calibration of Models A and B with QFracture in the held-out QResearch validation cohort and the external CPRD validation cohort. Overall discrimination was high and similar across models with small numerical increases for Model B in some comparisons. For example, C-index values for Model B for major osteoporotic fracture in women were 0.815 (0.813 to 0.817) in CPRD and 0.816 (0.814 to 0.818) in QResearch. In men, the corresponding values were 0.766 (0.761 to 0.771) and 0.784 (0.780 to 0.788) (Supplementary Table 1). Corresponding values for hip fracture in women were 0.889 (0.887 to 0.891) and 0.900 (0.898 to 0.902) and in men were 0.853 (0.847 to 0.859) and

0.885 (0.881 to 0.888). Discrimination varied across ethnic groups, although the pattern differed by sex and fracture outcome and confidence intervals were wide for some smaller groups (Supplementary Table 2). Condition-specific discrimination was generally similar between Model B, Model A and QFracture (Supplementary Tables 1–2) though estimates were imprecise for several uncommon conditions. Within age bands, discrimination was more moderate and declined at older ages.

Overall calibration accounting for competing risks was substantially better for Models A and B compared with QFracture in both validation cohorts (Table 2, Figure 4) while subgroup estimates were heterogeneous and sometimes imprecise (Supplementary Tables 3 & 4). For major osteoporotic fracture observed-to-expected ratios of 0.950–1.122 for Model A and B indicated a degree of under- or over-prediction depending on sex and cohort, with calibration intercepts close to zero and calibration slopes of 0.918–1.074. Predicted and observed risks agreed closely across most of the risk distribution for models A and B, although the plots indicated some under-prediction of hip fracture in women in CPRD and some overprediction among individuals at the highest predicted risks. QFracture overpredicted risk for both outcomes in men and women: observed-to-expected ratios ranged from 0.499 to 0.758, with increasing divergence at higher predicted risks, negative calibration intercepts and attenuated calibration slopes.

Models A and B showed threshold-dependent net benefit gains over QFracture on decision-curve analysis, while their curves were generally close (Supplementary Figure 3). Differences were most apparent among women and for hip fracture prediction and were smaller in men and in the CPRD validation cohort. For example, at the 10% threshold the net benefit per 10,000 women for major osteoporotic fracture was 60.8 for model A, 61.6 for model B, and 45.9 for QFracture.

Reclassification comparisons

Sensitivity and specificity of all three models were similar in equal-sized high-risk group comparisons in the held-out QResearch cohort. The equal sized groups correspond to the numbers classified as high risk by QFracture at a 10-year predicted risk of $\geq 10\%$ for major osteoporotic fracture ($n=367,297$) and $\geq 3\%$ for hip fracture using QFracture, (Table 3). Characteristics of participants at high risk with each model and those reclassified using Model B compared with QFracture are shown in Supplementary Tables 6 (major osteoporotic fracture) and 7 (hip fracture). Those reclassified up by Model B compared with QFracture were younger, more likely to be women and generally had a higher prevalence of the newly incorporated predictors.

The observed 10-year risk of major osteoporotic fracture among participants classified up with Model B compared with QFracture was 7.93% (95% CI 7.58 to 8.30) whereas the observed risk among those classified down with Model B was 6.50% (6.17 to 6.83). The corresponding figures for hip fracture were 3.30% (2.98 to 3.64) and 2.29% (2.00 to 2.56).

Omitting ethnicity from the prediction models produced similar overall discrimination: C-indices for Model B versus Model C were 0.816 versus 0.812 in women and 0.784 versus

0.778 in men for major osteoporotic fracture, and 0.900 versus 0.898 and 0.885 versus 0.882, respectively, for hip fracture (Supplementary Tables 8 and 9). However, calibration changed materially in the held-out QResearch cohort: Model B showed modest overall underprediction (observed-to-expected ratios 1.079–1.122), whereas Model C overpredicted risk (0.920–0.951). Across the non-White ethnic groups examined, Model C consistently produced lower observed-to-expected ratios than Model B, whose estimates were generally closer to one, although confidence intervals were wide in some smaller groups (Extended Data Figure 5 and Supplementary Figures 4–5).

Brier Scores

The absolute 10-year Brier scores for Models A and B were similar overall and within age bands for both fracture outcomes in women and men (Supplementary Table 5). Overall Model B scores were 0.0377 for major osteoporotic fracture and 0.0178 for hip fracture in women, and 0.0171 and 0.0091, respectively, in men.

In paired analyses of 10-year Brier scores, Model B and Model A had similar prediction error across most age and ethnic groups, although point estimates increasingly favoured Model B at older ages (Supplementary Figures 6–9). Model B generally had lower (more favourable) Brier scores than Model C, which omitted ethnicity, with the largest differences observed at older ages and for major osteoporotic fracture. Improvements over the original QFracture algorithm were substantially larger and increased markedly among participants aged 80–89 and 90 years or older in both women and men.

DISCUSSION

In this population-based study of more than 15 million adults, we developed and validated contemporary models for 10-year major osteoporotic fracture and hip fracture that explicitly represent several populations and treatment exposures not included in the core predictor sets of commonly used tools. A principal clinical and conceptual advance of Model B is the simultaneous inclusion, within a population-based model, of learning disability, Down syndrome, multiple sclerosis, visual impairment, HIV/AIDS, selected cancers and cancer-directed endocrine therapies, while also accounting explicitly for competing mortality. These characteristics are increasingly relevant as survival with cancer, HIV/AIDS and complex chronic disease improves, yet they are not explicitly represented in QFracture or the FRAX models^{4,21}. It is particularly relevant to the recently published UK NICE guideline on osteoporosis risk assessment, which retains these tools while emphasising the need to consider additional clinical risk factors not captured within the assessment⁴. Their addition produced several substantial, age-dependent prognostic associations and shifted model-based absolute-risk trajectories towards illustrative risk thresholds at younger ages than in otherwise comparable individuals. These findings identify a potential targeted role for the additional predictors in prompting fracture-risk assessment before some individuals would ordinarily be identified through age and conventional osteoporosis risk factors alone. Earlier

identification could prompt timely review of bone health, falls risk, medication, nutrition and the need for bone-density testing.

Cancer survival has improved leading to a growing population exposed to the skeletal effects of malignancy, reduced activity and prolonged treatment.⁷ Aromatase inhibitors used in post-menopausal women with breast cancer and androgen-deprivation therapy used in men with prostate cancer are established threats to bone health.^{11,22} In men with non-metastatic prostate cancer, rapid bone loss occurs within the first year of starting androgen deprivation therapy and may affect up to 80% of men within 10 years of treatment.¹¹ International guidance recommends regular fracture-risk assessment for men receiving androgen-deprivation therapy.^{23,24} However, this exposure is not explicitly incorporated into currently available general-population risk tools, and fewer than 10% of eligible men may undergo bone-density assessment.²⁵ Similarly, in women with hormone-receptor-positive breast cancer, aromatase-inhibitor therapy can accelerate bone loss, potentially compounding the skeletal effects of chemotherapy-induced ovarian failure and other cancer treatment.¹¹ Guidelines recommend assessment of fracture risk in such patients despite this not being factored into risk assessment tools.⁴ At the reference ages used in Figure 2, aromatase-inhibitor treatment in women and androgen-deprivation therapy in men were associated with approximately 2-fold and 3.6-fold adjusted hazards of hip fracture, respectively; the magnitude of these associations varied with age where interaction terms were included. Explicit incorporation into general-population fracture-risk prediction, therefore, may help identify cancer patients who would benefit from osteoporotic risk estimation and clinical discussions about assessment and prevention.

People living with HIV have lower bone mineral density and greater fracture risk, particularly during periods of active disease and following initiation of antiretroviral therapy.⁸ ²⁶ These clinical contacts provide opportunities for bone-health assessment. In our analysis, Model B, HIV/AIDS was retained as an additional predictor of hip fracture in men and was associated with a 43% higher adjusted hazard of hip fracture (HR 1.43, 95% CI 1.06 to 1.92). Explicit inclusion of HIV in an electronic health-record-based prediction model could therefore reduce reliance on clinicians recognising and separately accounting for a risk factor omitted from existing tools.

The implications may be particularly important for people with learning disability or Down syndrome, who may experience both increased fracture risk and barriers to conventional osteoporosis assessment.²⁷ Our study extends the earlier IDFracture work¹⁰ by evaluating these characteristics within a general-population model, using a more contemporary study period, a wider age range, more complete recording of ethnicity and alcohol consumption, and explicit accounting for competing mortality. In the illustrative trajectories, otherwise comparable people with learning disability reached a 10% predicted MOF risk approximately 10 years earlier than those without learning disability, while men with Down syndrome reached a 3% hip-fracture risk approximately 20 years earlier. These estimates do not themselves establish intervention thresholds, but indicate how explicit representation could prompt earlier review rather than relying on age or conventional risk factors alone.

Our findings also address an important limitation of QFracture by estimating cumulative fracture risk in the presence of competing mortality.⁴ Compared with QFracture, Model A showed substantially better calibration and lower Brier scores, indicating reduction in prediction error particularly at older ages. Accurate calibration is particularly important where absolute predicted risk informs subsequent investigation or intervention because systematic over- or underprediction can alter which individuals cross clinically relevant risk thresholds. Because the Brier score evaluates prediction error across the full probability distribution, these findings complement the threshold-specific decision-curve results: together they indicate improved predictive performance and estimated clinical utility.

We utilised illustrative thresholds based on examples used in international guidance including 3% for hip fracture, commonly used in the United States³, New Zealand⁶ and Australia⁵ and 10% for major osteoporotic fracture used in the UK.⁴ In equal-sized high-risk groups, observed fracture risk was higher among participants reclassified upwards by Model B compared with QFracture than among those reclassified downwards. Nevertheless, the models generate continuous absolute-risk estimates and are not tied to a single intervention threshold. Appropriate thresholds will depend on whether the intended action is clinical review, bone-density testing or preventive treatment, and should reflect treatment benefits and harms, patient preferences, resource use, cost-effectiveness and national care pathways

The ethnicity-omission analysis illustrates the limitations of attempting to achieve fairness simply by removing a protected characteristic as recommended in a recent review of the United States version of FRAX.³ Discrimination measures within ethnic groups changed little after ethnicity was omitted, whereas calibration generally deteriorated, producing greater overprediction among the non-White groups examined. This could lead to unnecessary treatment with potential side effects. These findings illustrate that simply omitting a protected characteristic does not guarantee equitable predictions and that removal is itself an active modelling choice. They do not establish that including ethnicity makes a model fair; implementation would require ongoing evaluation of performance and clinical consequences across population groups. The models excluding ethnicity may however be more appropriate for use in countries where the ethnic groups differ from those in the UK or where differences in fracture risk between ethnic groups may vary from those in the UK.

The study has several strengths. Its scale yielded precise overall estimates and enabled evaluation of risks for uncommon but clinically important conditions that smaller general-population cohorts cannot examine reliably. The nested comparison of Model A and Model B helped distinguish improvements arising from contemporary re-estimation and competing-risk modelling from those attributable to the additional predictors. The comparison between Model B and Model C specifically highlights the contribution of including ethnicity to model performance. QResearch practices were separated for development and held-out validation, reducing leakage between these stages, and validation in CPRD tested performance in a different primary-care database, clinical software system and time period. Linkage to hospital-admission and mortality records strengthened ascertainment of fractures and

competing deaths. The analysis considered two clinically important outcomes, modelled women and men separately, allowed non-linear and age-dependent associations, addressed missing data using multiple imputation and evaluated discrimination, calibration, decision curves and reclassification. Our approach could also be replicated using linked electronic health records in other countries, enabling locally developed and validated models that reflect country-specific fracture incidence, competing mortality, population composition and clinical practice.

Several limitations warrant consideration. Diagnoses, prescriptions and fractures were derived from routinely recorded data and were not formally adjudicated; coding completeness may vary between practices, databases and patient groups. The cancer predictors in this study were based on recorded diagnoses and did not distinguish active from historical disease or specific treatment histories. Overall discrimination partly reflected age and was more moderate within age bands. Although the size of the cohort enabled analysis of uncommon conditions, subgroup estimates for some conditions and ethnic groups remained imprecise. The relatively small changes in population-average measures such as the C-index and Brier score should be interpreted in light of the low prevalence of these predictors. Such measures are dominated by common predictors—particularly age—and by the much larger proportion of participants without the additional conditions, making them relatively insensitive to clinically meaningful improvements concentrated in small subgroups. CPRD provided database-distinct validation in a temporally non-overlapping but earlier historical period than QResearch. Missing ethnicity data in CPRD precluded meaningful external evaluation across ethnic groups. Direct individual-level comparison with FRAX was not possible because its complete equations are not publicly available. Direct comparison with CFracture was not undertaken because there is no publicly available implementation²⁸. However, the C-index values for Models A and B in the external CPRD validation cohort were similar to those reported for CFracture in a held-out CPRD validation dataset, both overall and across age groups²⁹.

In conclusion, contemporary estimation within a competing-risk framework produced the largest improvement in population-level calibration and estimated clinical utility, whereas the addition of predictors for learning disability, multiple sclerosis, HIV/AIDS, selected cancers and cancer-directed endocrine therapies provided the principal clinical and conceptual novelty. Model B offers broader explicit clinical representation and reclassified risk at illustrative thresholds for individuals, often at younger ages, whose conditions are not explicitly represented in existing tools. Its effect on assessment, treatment and outcomes can now be evaluated and potentially recalibrated in settings beyond the UK. More broadly, these findings support prediction models that evolve as survival, treatments and the populations living with chronic disease change, rather than remaining fixed to historical cohorts and predictor sets.

ACKNOWLEDGMENTS

We acknowledge the contribution of EMIS practices who contribute to QResearch and Queen Mary University of London for expertise in developing and supporting the QResearch database. We acknowledge the contribution of patients and professionals who reviewed this application when it was submitted to the QResearch scientific committee including the research question, study design and lay summary. This project involves data derived from patient-level information collected by the NHS, as part of the care and support. The Hospital Episode Statistics data, and civil registration mortality data are used by permission from NHS England who retain the copyright in that data. NHS England bears no responsibility for the analysis or interpretation of the data.

AUTHOR CONTRIBUTIONS STATEMENT

JHC initiated the study, development of the research question, undertook the literature review, data extraction, data manipulation and primary data analysis and wrote the first draft of the paper. CC contributed to the refinement of the research question, design, analysis, interpretation and drafting of the paper. JHC undertook final editing of the paper. Both authors commented critically on drafts of the manuscript and approved the final submission.

COMPETING INTERESTS STATEMENT

The authors declare the following competing interests: JHC reports grants from National Institute for Health Research, and other research councils, during the conduct of the study. JHC is director of QResearch, a not-for-profit unit based at Queen Mary University of London. Until 2023, JHC had a 50% shareholding in ClinRisk Ltd, co-owning it with her husband, who was an executive director. In 2023, 100% of the share capital was donated to Endeavour Health Care Charitable Trust and the company renamed to Endeavour Predict Ltd. JHC is a consultant to Endeavour Predict Ltd. and her husband is a non-executive director. The company licences software (including QFracture) both to the private sector and to NHS bodies or bodies that provide services to the NHS (through GP electronic health record providers, pharmacies, hospital providers and other NHS providers). CC reports receiving personal fees from ClinRisk Ltd until 2023, outside this work.

FUNDING: There was no external funding.

Table 1: Baseline characteristics of participants. Baseline characteristics are shown for individuals aged 30-100 years in the QResearch derivation (England) and held-out validation cohorts (England only) and in the external CPRD validation cohort (England and Wales). Values are numbers (%) of participants unless indicated otherwise.

	<i>QResearch Derivation cohort</i>	<i>QResearch held-out validation cohort</i>	<i>CPRD external validation cohort</i>
Total participants	9,465,613	3,155,655	3,011,531
Female	4716370 (49.8)	1571806 (49.8)	1538596 (51.1)
Male	4749243 (50.2)	1583849 (50.2)	1472935 (48.9)
Mean age (SD)	51.3 (16.6)	51.2 (16.5)	51.6 (16.2)
Body mass index (BMI) recorded	7201765 (76.1)	2400217 (76.1)	2128340 (70.7)
Mean BMI (SD)	26.8 (5.4)	26.8 (5.4)	26.1 (4.5)
Ethnicity recorded	8860019 (93.6)	2950783 (93.5)	817964 (27.2)
White	7002950 (74.0)	2289950 (72.6)	689518 (22.9)
Indian	350765 (3.7)	122760 (3.9)	26513 (0.9)
Pakistani	182217 (1.9)	63199 (2.0)	10141 (0.3)
Bangladeshi	113397 (1.2)	33826 (1.1)	3341 (0.1)
Other Asian	232969 (2.5)	77718 (2.5)	18473 (0.6)
Caribbean	102194 (1.1)	39434 (1.2)	11290 (0.4)
Black African	293415 (3.1)	123090 (3.9)	23006 (0.8)
Chinese	96637 (1.0)	33890 (1.1)	5887 (0.2)
Other	485475 (5.1)	166916 (5.3)	29795 (1.0)
Smoking recorded	8990523 (95.0)	3005182 (95.2)	2767130 (91.9)
Non-smoker	5136744 (54.3)	1714571 (54.3)	1550441 (51.5)
Ex smoker	2094868 (22.1)	695253 (22.0)	493325 (16.4)
Light smoker (<10/day)	996738 (10.5)	332240 (10.5)	181539 (6.0)
Moderate smoker (10-19/day)	496467 (5.2)	170712 (5.4)	326785 (10.9)
Heavy smoker (≥20/day)	265706 (2.8)	92406 (2.9)	215040 (7.1)
Alcohol recorded	6284565 (66.4)	2214324 (70.2)	2401138 (79.7)
Non drinker	2534020 (26.8)	894478 (28.3)	446665 (14.8)
Trivial <1u/day	1911984 (20.2)	675891 (21.4)	1073478 (35.6)
Light 1-2u/day	1031705 (10.9)	359129 (11.4)	642765 (21.3)
Moderate 3-6u/day	672290 (7.1)	236124 (7.5)	194854 (6.5)
Heavy 7-9u/day	61241 (0.6)	21374 (0.7)	22690 (0.8)
Very Heavy >9u/day	73325 (0.8)	27328 (0.9)	20686 (0.7)
Family history of osteoporosis	51746 (0.5)	15260 (0.5)	11935 (0.4)
Corticosteroids	440124 (4.6)	143918 (4.6)	102306 (3.4)
Antidepressants	1502478 (15.9)	491696 (15.6)	334614 (11.1)
Oestrogen HRT	128621 (1.4)	42734 (1.4)	121158 (4.0)
Asthma/COPD	1161460 (12.3)	385244 (12.2)	49127 (1.6)
Any cancer	402307 (4.3)	130362 (4.1)	79746 (2.6)
Cardiovascular disease	609191 (6.4)	198987 (6.3)	204749 (6.8)
Dementia	170564 (1.8)	54603 (1.7)	24671 (0.8)

Falls	564345 (6.0)	182087 (5.8)	102998 (3.4)
Chronic liver disease	99286 (1.0)	33080 (1.0)	7528 (0.2)
Parkinson's disease	36122 (0.4)	11747 (0.4)	10100 (0.3)
Rheumatoid or SLE	193494 (2.0)	62716 (2.0)	26390 (0.9)
Chronic kidney disease	440495 (4.7)	141788 (4.5)	26616 (0.9)
Type 1 diabetes	41010 (0.4)	13720 (0.4)	9512 (0.3)
Type 2 diabetes	596038 (6.3)	198116 (6.3)	108577 (3.6)
Endocrine disorders	77021 (0.8)	25061 (0.8)	19644 (0.7)
Malabsorption	136057 (1.4)	45344 (1.4)	26806 (0.9)
Care home resident	270559 (2.9)	77857 (2.5)	9273 (0.3)
Epilepsy diagnosis or treatment	393171 (4.2)	128958 (4.1)	47748 (1.6)
Prior osteoporotic fracture	315042 (3.3)	102843 (3.3)	55264 (1.8)
New predictors in final models^y			
Learning disability	33556 (0.4)	10052 (0.3)	181 (0.0)
Down syndrome	3901 (0.0)	1218 (0.0)	1456 (0.0)
Heart Failure	125581 (1.3)	41470 (1.3)	37703 (1.3)
Multiple Sclerosis	23344 (0.2)	7724 (0.2)	7581 (0.3)
Visual impairment	122694 (1.3)	39536 (1.3)	34724 (1.2)
HIV/AIDS	21030 (0.2)	7969 (0.3)	1997 (0.1)
Aromatase inhibitors	24952 (0.3)	8094 (0.3)	2694 (0.1)
Androgen deprivation therapy	22361 (0.2)	7031 (0.2)	5613 (0.2)
Prostate cancer	62880 (0.7)	19913 (0.6)	8670 (0.3)
Blood cancer ^f	48217 (0.5)	15887 (0.5)	10438 (0.3)
Pancreatic cancer	2306 (0.0)	776 (0.0)	315 (0.0)
Cervical cancer	8159 (0.1)	2773 (0.1)	2510 (0.1)
Colorectal cancer	43170 (0.5)	13876 (0.4)	8977 (0.3)
Liver cancer	926 (0.0)	323 (0.0)	75 (0.0)
Lung cancer	15014 (0.2)	5040 (0.2)	2245 (0.1)
New predictors not in final models			
Renal cancer	27653 (0.3)	8928 (0.3)	5540 (0.2)
Uterine cancer	11789 (0.1)	3904 (0.1)	2387 (0.1)
Ovarian cancer	7963 (0.1)	2563 (0.1)	1879 (0.1)
Breast cancer	109008 (1.2)	35260 (1.1)	24903 (0.8)
Proton pump inhibitors	1828863 (19.3)	597240 (18.9)	240046 (8.0)
Hearing loss	707099 (7.5)	224433 (7.1)	138472 (4.6)
Cystic fibrosis	1956 (0.0)	609 (0.0)	2188 (0.1)

Definitions: All predictors were assessed at cohort entry and treated as fixed baseline covariates. Medication exposure was defined as at least two prescriptions before cohort entry, including one within the preceding 28 days. Diagnoses were defined as ever recorded before or on cohort entry. Valid BMI ranges were 15-50 kg/m². Cancer predictors were based on recorded diagnoses and did not distinguish active from historical disease, metastatic status, stage or severity.

^y New predictors included in one or more of final model B

^f Blood cancer includes acute and chronic leukaemia, Hodgkin's and non-Hodgkin's lymphoma and myeloma.

Table 2: Discrimination and calibration of the fracture prediction algorithms in the held-out QResearch validation cohort (England only 2013-2024) and external CPRD validation cohort (England and Wales, 2001-2012). Performance was evaluated in people aged 30-100 years for the major osteoporotic fracture and hip fracture outcome. QFracture is the original model from 2016. Model A includes the same predictors as QFracture but accounts for competing risks. Model B is based on Model A with additional predictors.

	<i>Validation cohort</i>	<i>Model</i>	<i>O/E ratio (95% CI)</i>	<i>Calibration Intercept (95% CI)</i>	<i>Calibration Slope (95% CI)</i>	<i>C-index (95% CI)</i>
Major osteoporotic fracture in women	QResearch	Model A	1.120 (1.108 to 1.132)	0.044 (0.031 to 0.057)	0.918 (0.904 to 0.933)	0.815 (0.813 to 0.817)
		Model B	1.122 (1.110 to 1.134)	0.045 (0.032 to 0.059)	0.930 (0.915 to 0.945)	0.816 (0.814 to 0.818)
		QFracture	0.653 (0.646 to 0.660)	-0.969 (-0.987 to -0.952)	0.567 (0.561 to 0.573)	0.814 (0.812 to 0.816)
	CPRD	Model A	1.055 (1.044 to 1.065)	0.050 (0.038 to 0.063)	1.034 (1.018 to 1.049)	0.815 (0.812 to 0.817)
		Model B	1.063 (1.053 to 1.074)	0.053 (0.041 to 0.065)	1.040 (1.024 to 1.056)	0.815 (0.813 to 0.817)
		QFracture	0.698 (0.691 to 0.705)	-0.597 (-0.611 to -0.583)	0.710 (0.703 to 0.718)	0.814 (0.812 to 0.816)
Major osteoporotic fracture in men	QResearch	Model A	1.087 (1.070 to 1.105)	0.068 (0.048 to 0.089)	0.957 (0.939 to 0.975)	0.783 (0.779 to 0.787)
		Model B	1.088 (1.071 to 1.106)	0.065 (0.045 to 0.086)	0.954 (0.936 to 0.973)	0.784 (0.780 to 0.788)
		QFracture	0.584 (0.575 to 0.594)	-1.767 (-1.806 to -1.727)	0.490 (0.483 to 0.497)	0.782 (0.778 to 0.786)
	CPRD	Model A	0.950 (0.933 to 0.967)	0.017 (-0.005 to 0.040)	1.074 (1.053 to 1.095)	0.765 (0.761 to 0.770)
		Model B	0.952 (0.936 to 0.969)	0.014 (-0.009 to 0.036)	1.067 (1.045 to 1.088)	0.766 (0.761 to 0.771)
		QFracture	0.654 (0.643 to 0.666)	-0.991 (-1.026 to -0.957)	0.628 (0.618 to 0.638)	0.766 (0.762 to 0.771)
Hip fracture in women	QResearch	Model A	1.075 (1.059 to 1.092)	-0.020 (-0.038 to -0.001)	0.847 (0.831 to 0.863)	0.899 (0.897 to 0.900)
		Model B	1.079 (1.063 to 1.096)	-0.017 (-0.035 to 0.002)	0.852 (0.835 to 0.869)	0.900 (0.898 to 0.902)
		QFracture	0.522 (0.514 to 0.531)	-1.443 (-1.468 to -1.417)	0.502 (0.496 to 0.508)	0.897 (0.895 to 0.899)
	CPRD	Model A	1.265 (1.248 to 1.282)	0.196 (0.180 to 0.212)	0.920 (0.903 to 0.938)	0.888 (0.881 to 0.894)
		Model B	1.285 (1.268 to 1.303)	0.187 (0.171 to 0.203)	0.886 (0.869 to 0.904)	0.889 (0.887 to 0.891)
		QFracture	0.748 (0.738 to 0.758)	-0.679 (-0.699 to -0.659)	0.607 (0.599 to 0.614)	0.888 (0.883 to 0.892)

Hip fracture in men	QResearch	Model A	1.098 (1.073 to 1.123)	0.021 (-0.007 to 0.048)	0.871 (0.850 to 0.892)	0.883 (0.880 to 0.887)
		Model B	1.095 (1.071 to 1.120)	0.006 (-0.022 to 0.034)	0.865 (0.843 to 0.886)	0.885 (0.881 to 0.888)
		QFracture	0.499 (0.488 to 0.511)	-2.241 (-2.298 to -2.185)	0.452 (0.445 to 0.458)	0.882 (0.879 to 0.886)
	CPRD	Model A	1.107 (1.082 to 1.133)	0.091 (0.063 to 0.120)	0.948 (0.923 to 0.973)	0.846 (0.841 to 0.850)
		Model B	1.132 (1.106 to 1.158)	-0.023 (-0.053 to 0.007)	0.817 (0.795 to 0.839)	0.853 (0.847 to 0.859)
		QFracture	0.712 (0.695 to 0.729)	-1.169 (-1.219 to -1.119)	0.560 (0.550 to 0.569)	0.845 (0.840 to 0.849)

Table 3 Model performance in exactly equal sized high-risk groups of men and women corresponding to a 10-year predicted risk of $\geq 10\%$ for major osteoporotic fracture (n=367,297) and $\geq 3\%$ for hip fracture using QFracture (n=497,738). Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were estimated using inverse-probability-of-censoring weights, with competing non-fracture deaths treated as known non-fracture outcomes. Values include 95% confidence intervals obtained from 1,000 paired non-parametric bootstrap samples. Analyses combine women and men.

<i>Outcome</i>	<i>Model</i>	<i>Model-specific cut-off</i>	<i>High risk, %</i>	<i>Sensitivity, % (95% CI)</i>	<i>Specificity, % (95% CI)</i>	<i>PPV, % (95% CI)</i>	<i>NPV, % (95% CI)</i>
Major osteoporotic fracture	QFracture	10.0%	12.0	51.79 (51.30 to 52.28)	86.70 (86.66 to 86.77)	10.63 (10.51 to 10.80)	98.33 (98.30 to 98.36)
	Model A	6.4%	12.0	52.39 (51.98 to 52.83)	86.48 (86.43 to 86.54)	10.58 (10.46 to 10.75)	98.35 (98.32 to 98.37)
	Model B	6.5%	12.0	52.57 (52.20 to 52.99)	86.49 (86.44 to 86.55)	10.62 (10.48 to 10.79)	98.35 (98.33 to 98.38)
Hip fracture	QFracture	3.0%	15.9	80.05 (79.47 to 80.71)	81.52 (81.46 to 81.60)	5.96 (5.87 to 6.05)	99.64 (99.63 to 99.66)
	Model A	2.2%	15.9	80.19 (79.61 to 80.72)	81.47 (81.41 to 81.55)	5.95 (5.86 to 6.04)	99.65 (99.63 to 99.66)
	Model B	2.2%	15.9	80.38 (79.83 to 80.92)	81.61 (81.55 to 81.68)	6.01 (5.91 to 6.10)	99.65 (99.64 to 99.66)

Figure 1: Final Model B adjusted hazard ratios for major osteoporotic fracture. Adjusted hazard ratios (HRs) for major osteoporotic fracture for men and women in the derivation cohort are shown. Estimates are presented at the mean age (52 years for women; 50 for men), for predictors modelled using age-interaction terms. Models also included ethnicity and were adjusted for fractional polynomial transformations of age and body mass index (BMI). Extended Data Figure 2 shows the corresponding age-dependent effects.

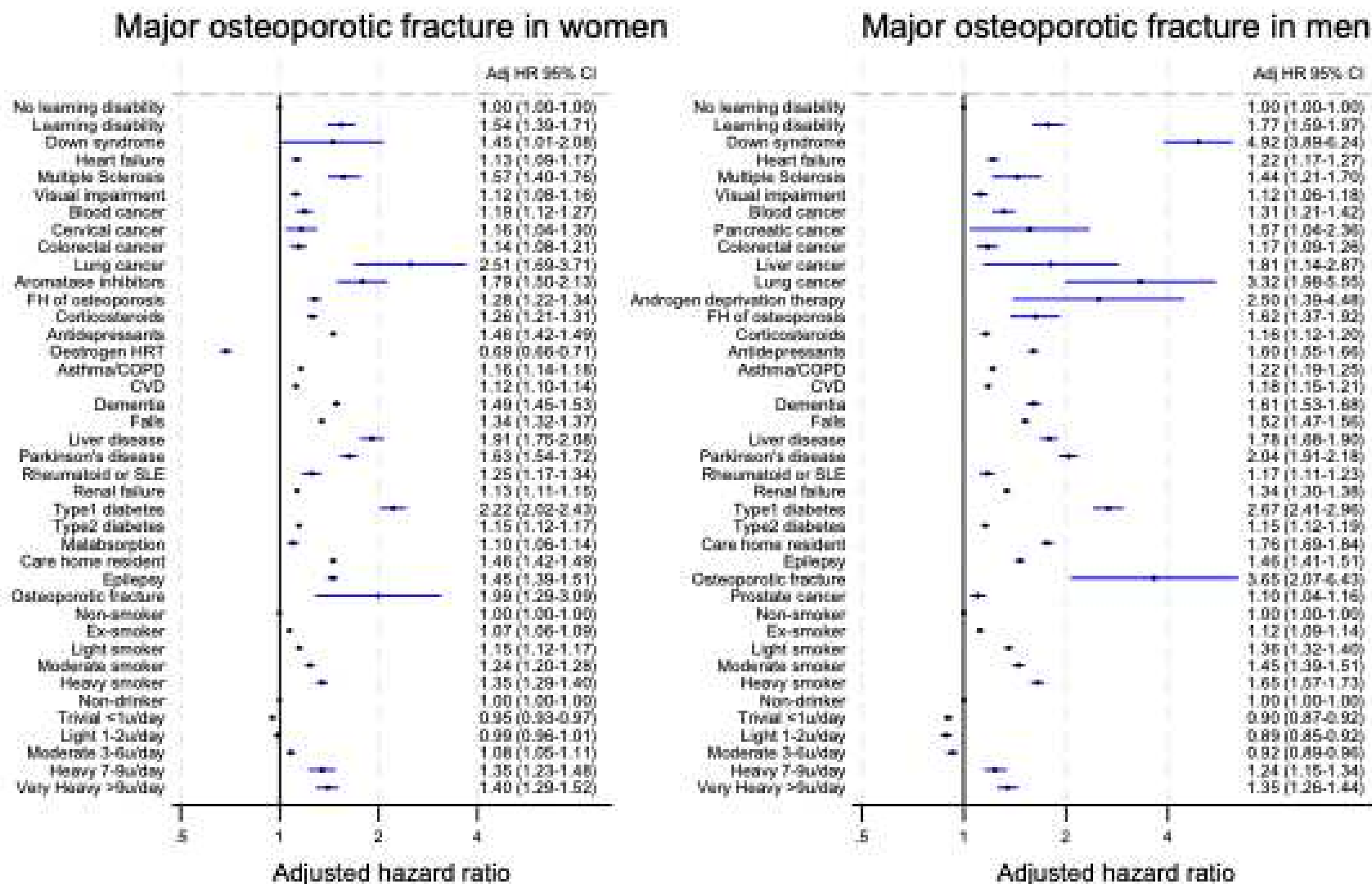


Figure 2: Final model B adjusted hazard ratios for hip fracture. Adjusted hazard ratios (HRs) for hip fracture among men and women in the derivation cohort are shown. Estimates are presented at the mean age (52 years for women; 50 for men) for predictors modelled using age-interaction terms. Models also included ethnicity and were adjusted for fractional polynomial transformations of age and BMI. Extended Data Figure 3 shows the corresponding age-dependent effects.

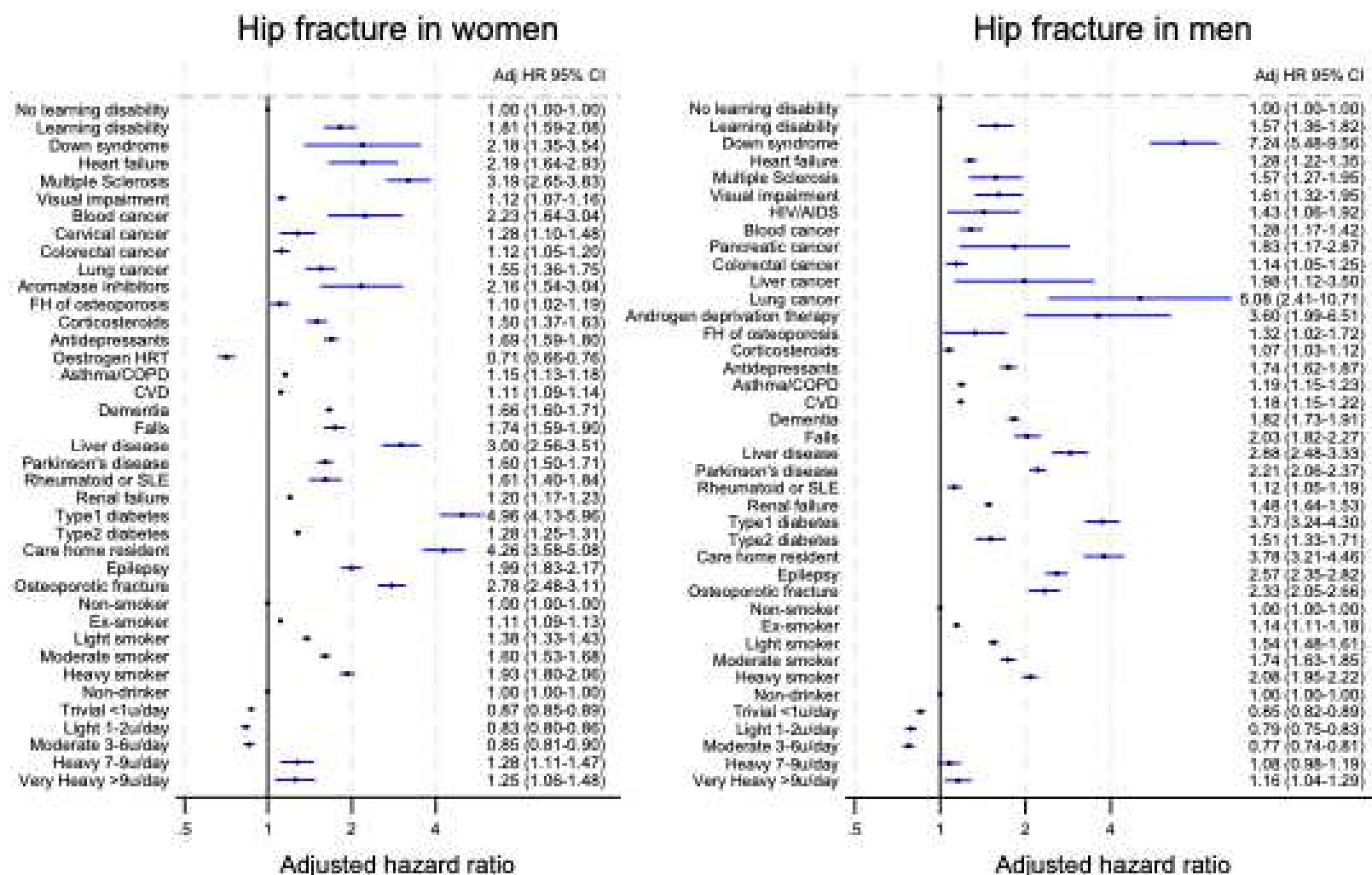


Figure 3: Effects of selected predictors on age-specific 10-year fracture risk. Predicted 10-year absolute risks of major osteoporotic fracture and hip fracture are shown for selected predictors retained in the final models according to age. Predictions were generated for otherwise comparable individuals with and without each predictor. The reference individual was of White ethnicity, had a body mass index of 25 kg/m², did not smoke or consume alcohol and had no additional adverse clinical characteristics. Horizontal lines indicate illustrative risk thresholds and should not be interpreted as universal treatment thresholds. The trajectories illustrate how several predictors shift fracture-risk trajectories towards higher values and result in clinically important threshold crossings at younger ages.

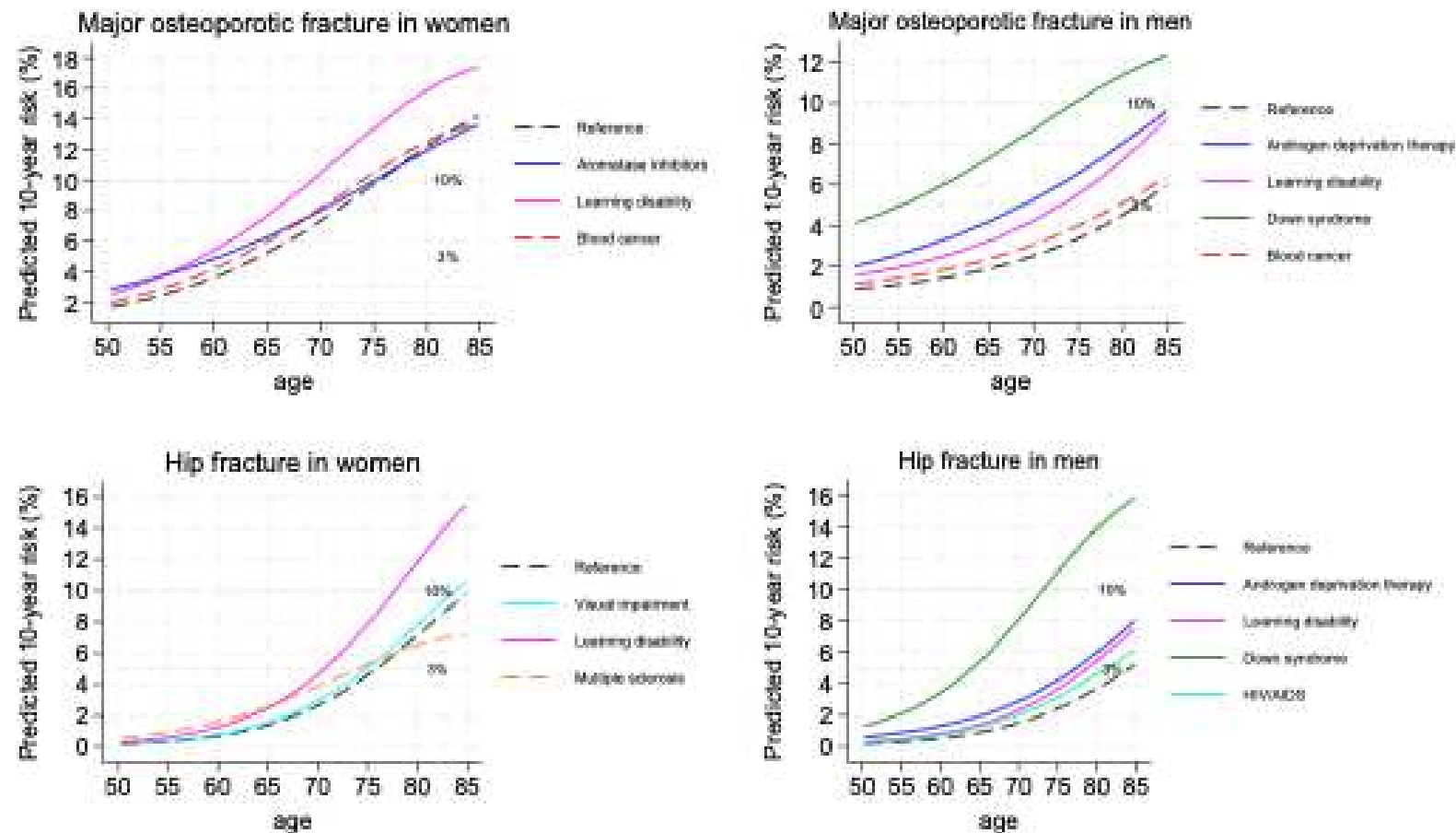
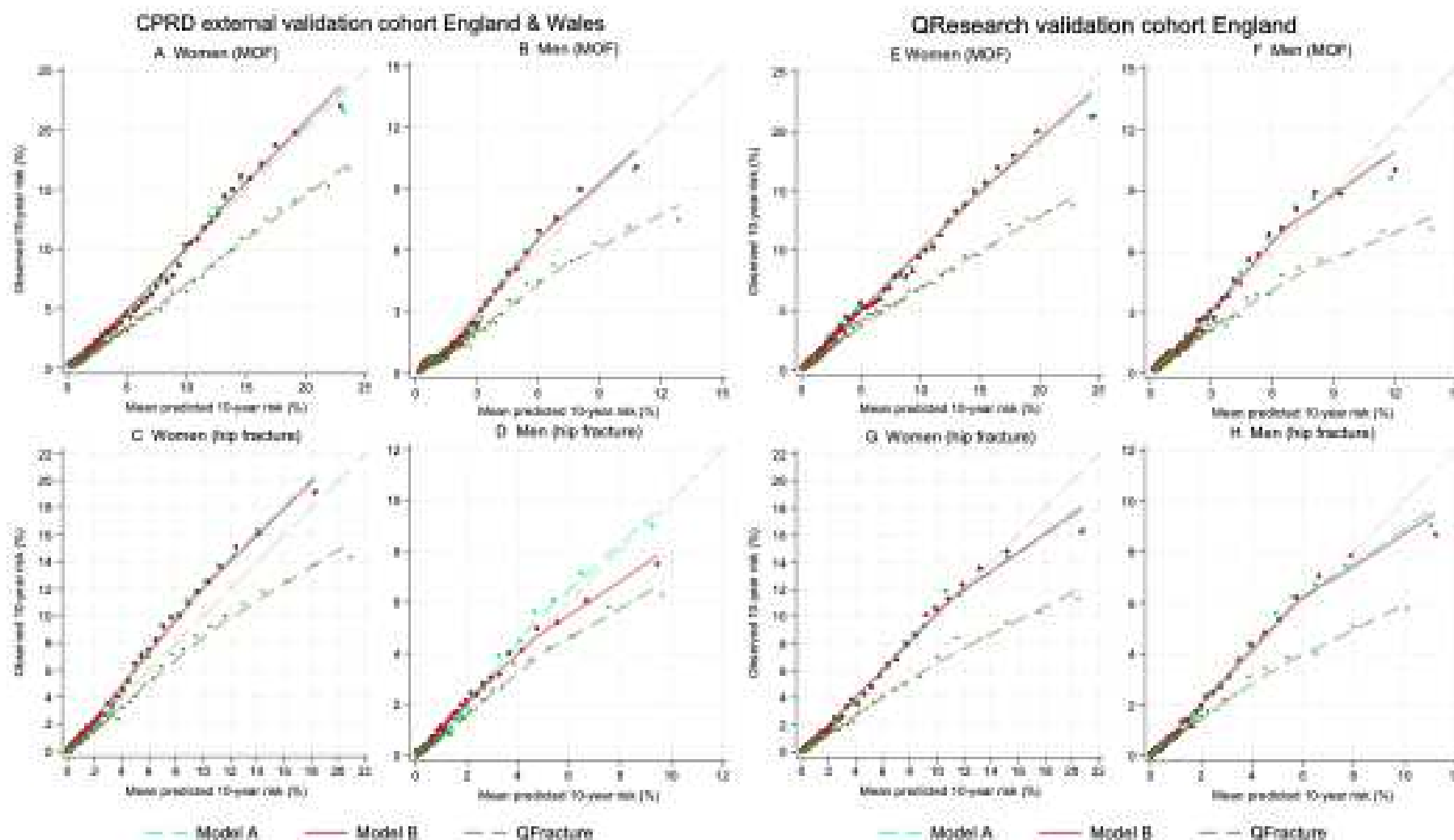


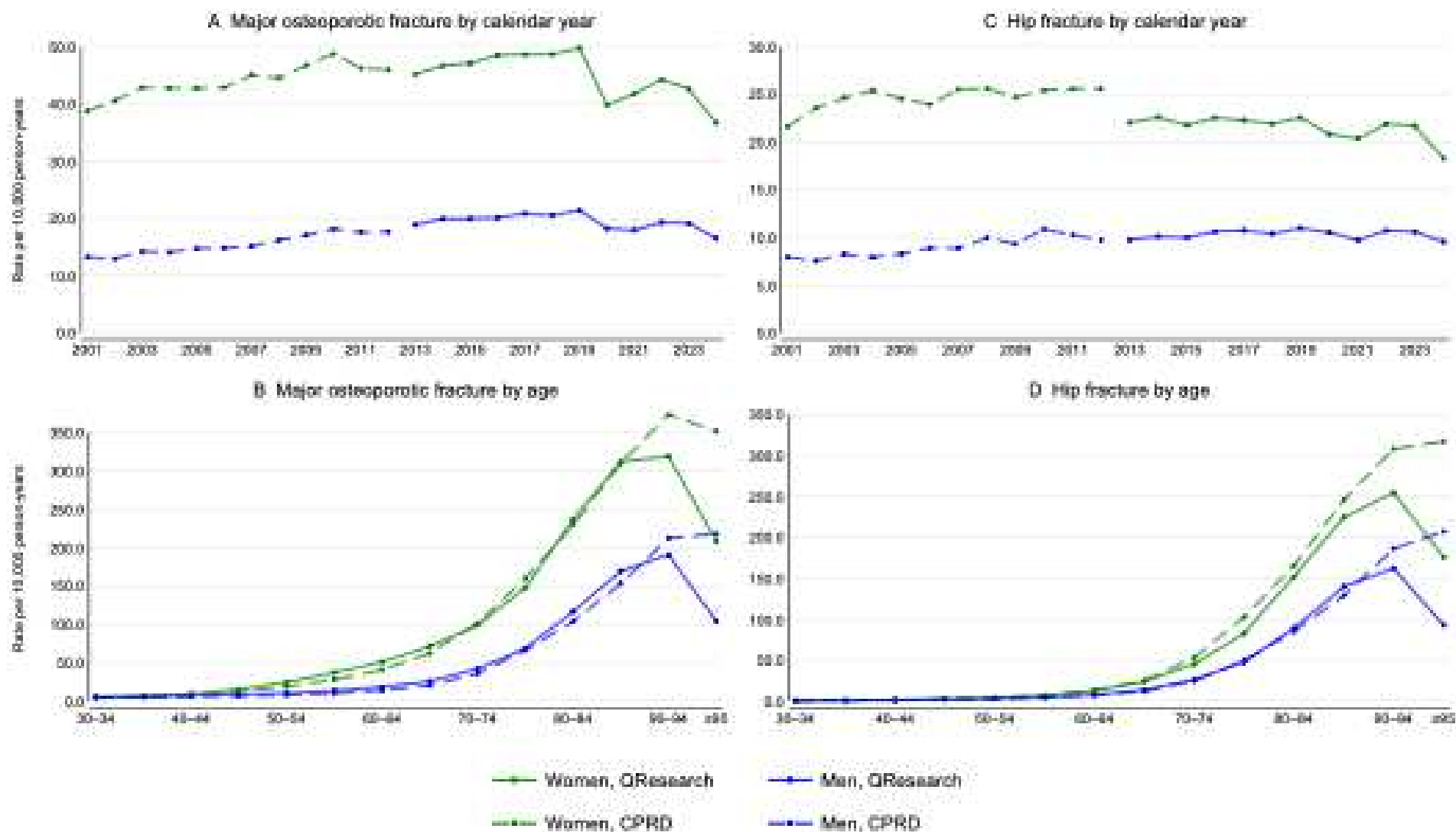
Figure 4: Calibration of Models A, Model B and QFracture in held-out QResearch and external CPRD validation cohorts. Centile calibration plots compare observed and predicted 10-year risks of major osteoporotic fracture and hip fracture in women and men. Points show observed cumulative incidence within centiles of predicted risk, accounting for competing non-fracture mortality; lines connect or smooth the corresponding centile estimates. The light-grey diagonal line indicates perfect calibration, where observed risk equals predicted risk. Plotting ranges were restricted to the displayed predicted-risk ranges to preserve visual resolution; complete calibration statistics are reported in Table 2.



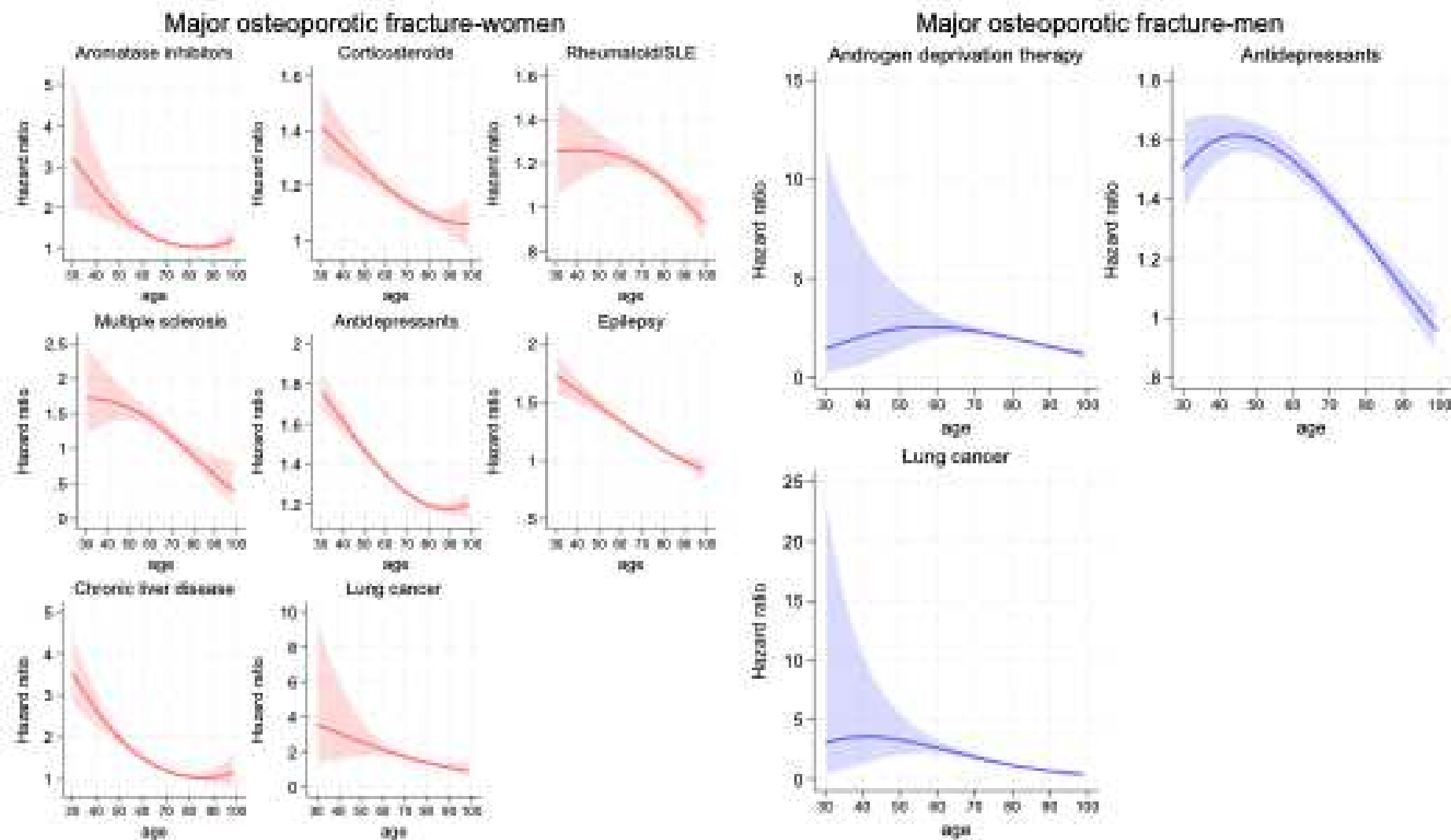
Extended Data Table 1: Flow of patients into the study for the QResearch derivation and validation cohort (2013-2024) and the CPRD validation cohort (2001-2012). Figures are counts or rates (95% Confidence Intervals) unless otherwise indicated.

	<i>QResearch Derivation</i>	<i>QResearch validation</i>	<i>CPRD validation</i>
Number of GP practices	1,047	348	382
Aged 30-100 years at cohort entry	9,465,613	3,155,655	3,011,531
Cohort follow up			
Number participants with >=10 years of follow up	2,622,775	923,912	948,504
Person years of follow-up (10,000)	5158	1750	1842
Major osteoporotic fracture			
Number of participants eligible for analysis	9,465,613	3,155,655	3,011,531
Cases during follow up	166,442	55,413	55,145
Cases within 10 years	157,119	52,192	51,435
Non-fracture deaths	620,403	201,368	239,241
Event rate per 10,000 person years	32.3 (32.1 to 32.4)	31.6 (31.4 to 31.9)	29.5 (29.2 to 29.7)
Death rate per 10,000 person years	120.3 (120.0 to 120.57)	115 (114.5 to 115.5)	129.9 (129.4 to 130.4)
Hip fracture			
Prior hip fracture at cohort entry excluded from cohort	93,938	30,583	16,707
Number of participants eligible for analysis	9,371,675	3,125,072	2,994,824
Cases during follow up	85,174	27,969	31,915
Cases within 10 years	80,203	26,257	29,726
Non-hip fracture deaths	620,286	201,364	239,235
Event rate per 10,000 person years	16.1 (15.9 to 16.2)	15.5 (15.4 to 15.7)	16.8 (16.6 to 16.9)

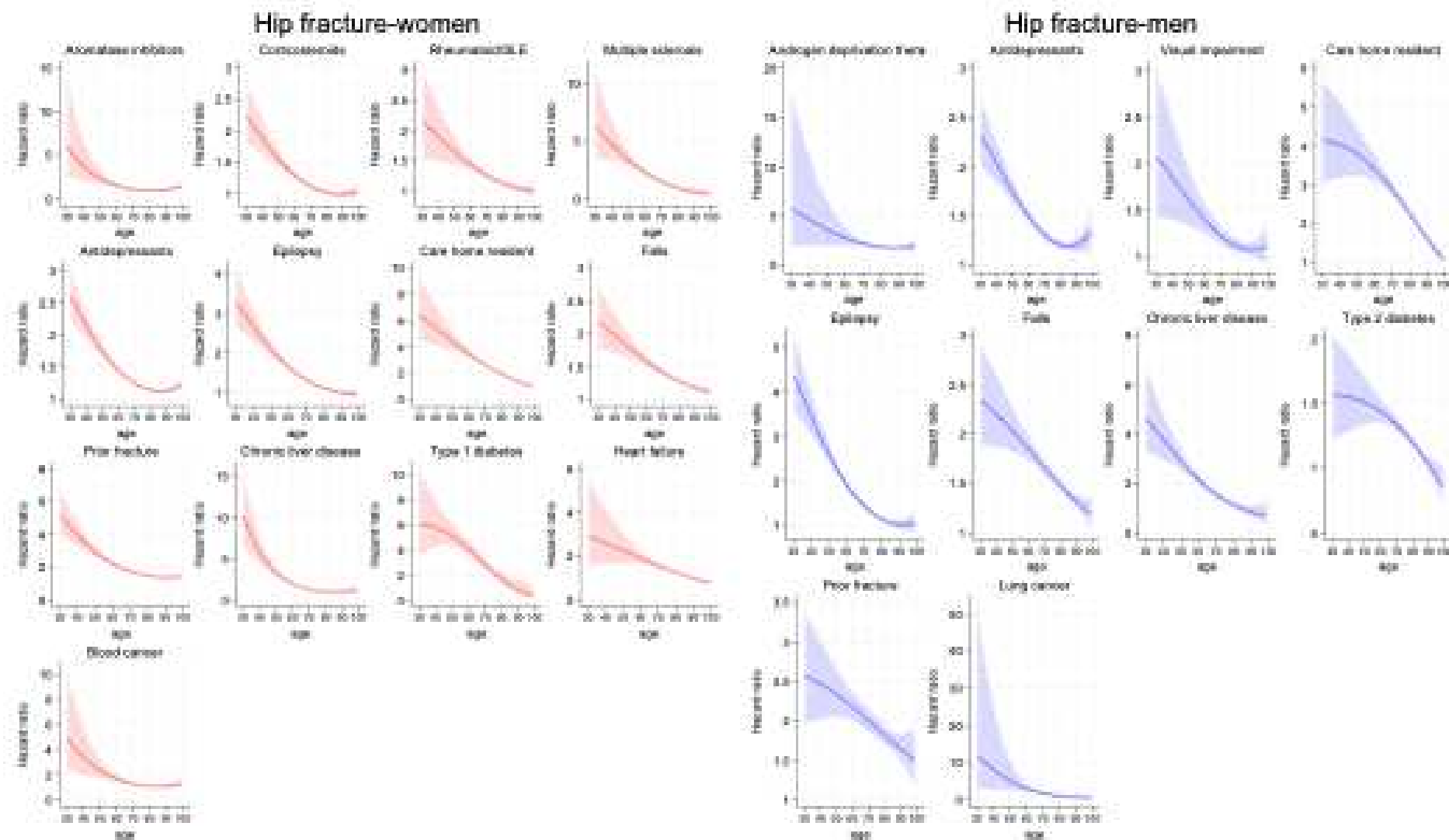
Extended Data Figure 1: Crude incidence rates of major osteoporotic fracture and hip fracture by calendar year and age. Crude incidence rates per 10,000 person-years are shown for major osteoporotic fracture and hip fracture according to calendar year (A, C) and age group (B, D) in women and men. Calendar-year analyses were restricted to 2001-2012 for CPRD and 2013-2024 for QResearch. Solid lines represent the held-out QResearch validation cohort and dashed lines represent the external CPRD validation cohort from England and Wales. Green lines denote women and blue lines denote men.



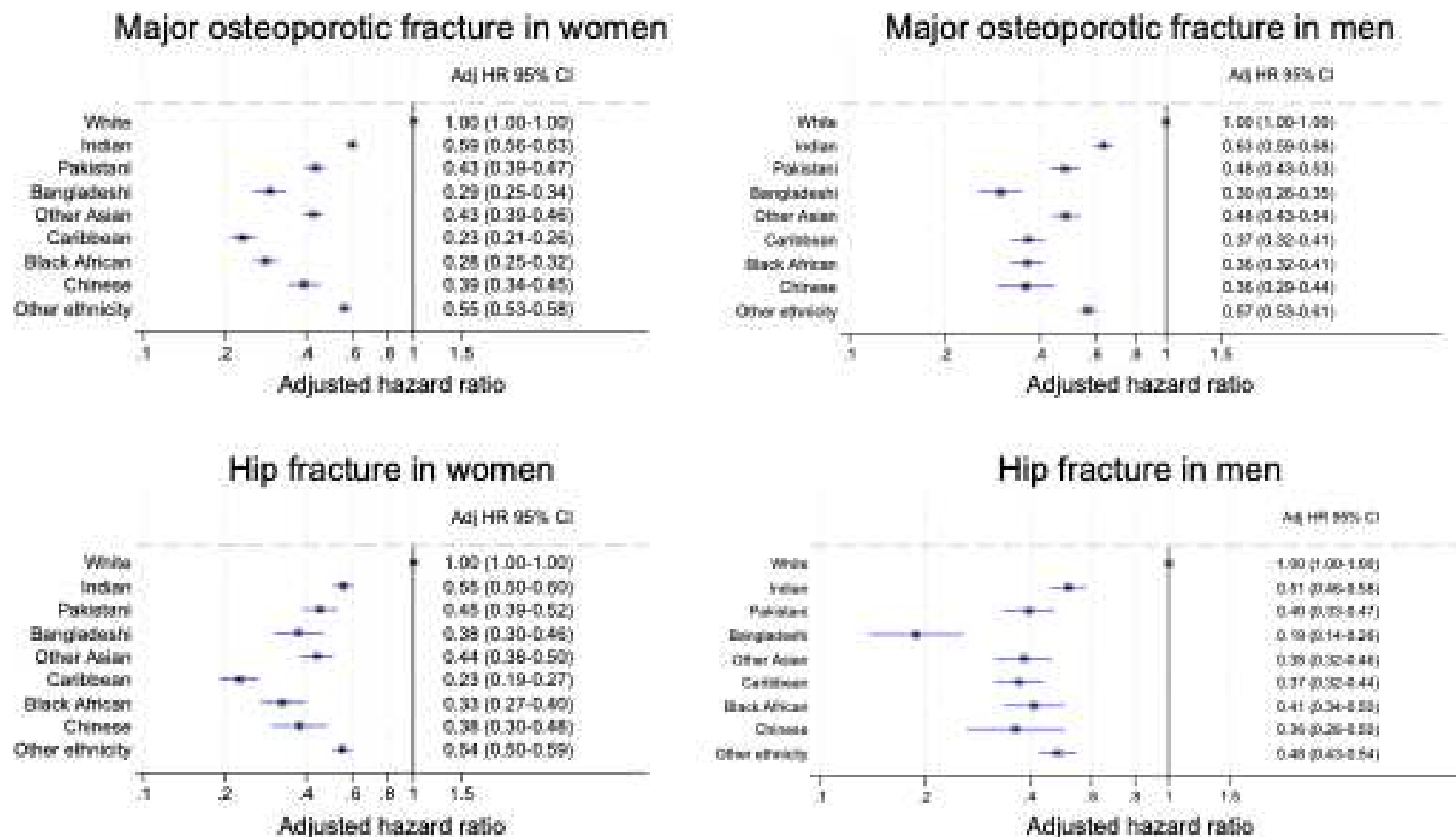
Extended Data Figure 2: Age-dependent effects of predictors included in the major osteoporotic fracture Model B. Estimated hazard ratios for predictors modelled using age-interaction terms are shown across the age range 30–100 years. Shaded areas indicate pointwise 95% confidence intervals. Y-axis scales differ between panels to facilitate visualisation of the age-dependent effects associated with individual predictors. The curves illustrate the changing magnitude of predictor effects across the life course.



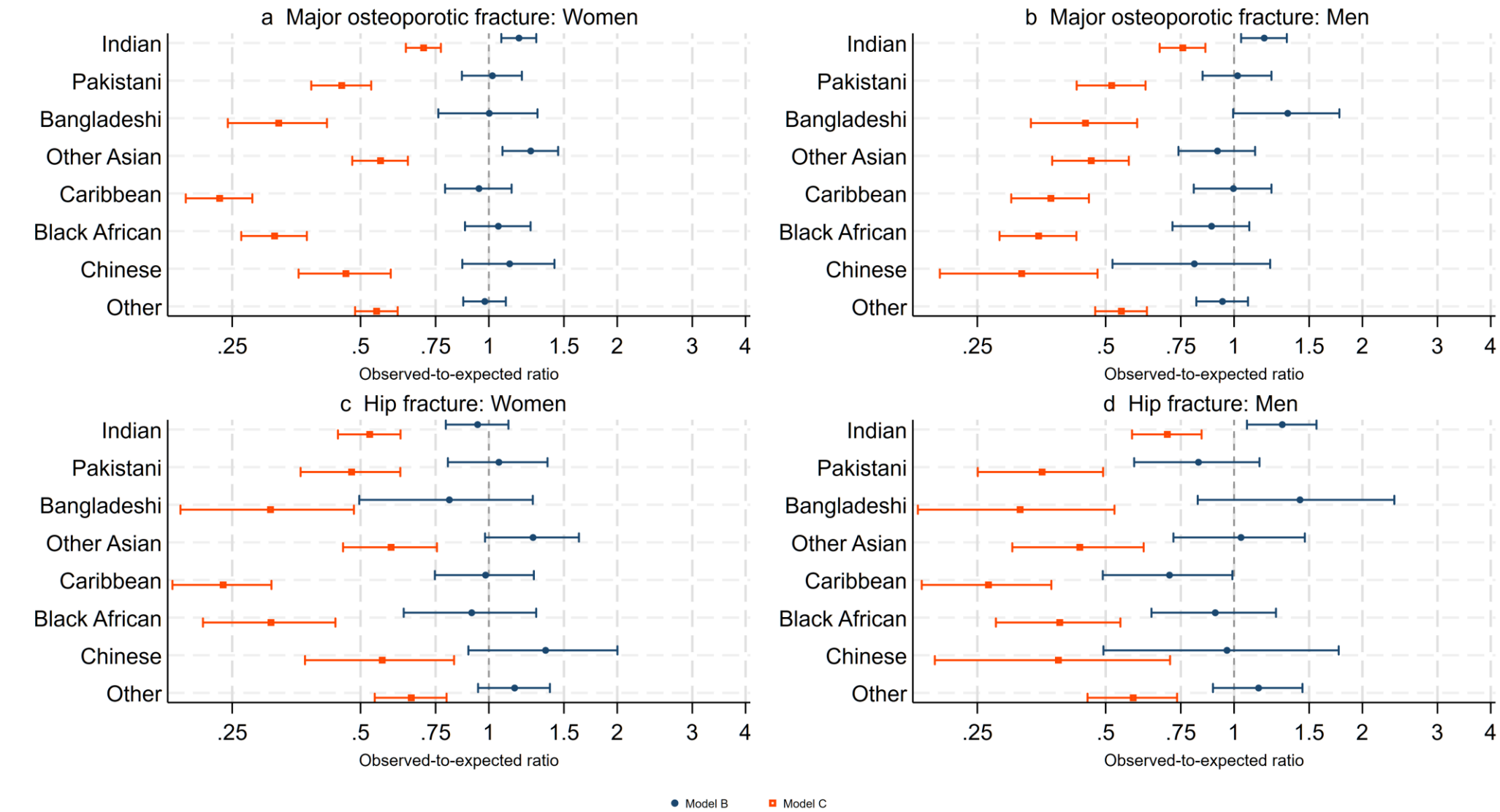
Extended Data Figure 3: Age-dependent effects of predictors included in the hip fracture Model B. Estimated hazard ratios for Model B predictors modelled using age-interaction terms are shown across the age range 30–100 years. Shaded areas indicate pointwise 95% confidence intervals. Y-axis scales differ between panels to facilitate visualisation of the age-dependent effects associated with individual predictors. The curves illustrate the changing magnitude of predictor effects across the life course.



Extended Data Figure 4: Final model B adjusted hazard ratios associated with ethnicity. Adjusted hazard ratios (HRs) with 95% CI for ethnicity for major osteoporotic fracture and HRs for hip fracture in the derivation cohort.



Extended Data Figure 5: Calibration of Model B and Model C in the held-out QResearch validation cohort with observed to expected ratio (95% confidence intervals) by ethnic group for major osteoporotic fracture and hip fracture in men and women.



REFERENCES

1. Collaborators, G.B.D.F. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet Healthy Longev* **2**, e580–e592 (2021).
2. Johannesdottir, F., Roberts, J.E., Kiel, D.P. & Tsai, J.N. Hip Fractures: A Review. *JAMA* **336**, 496–507 (2026).
3. LeBoff, M.S., *et al.* The clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int* **33**, 2049–2102 (2022).
4. National Institute for Clinical Excellence. NG259: Osteoporosis: assessing the risk of fragility fracture., Vol. 2026 (NICE, London, UK, 2026).
5. Royal Australian College of General Practitioners (RACGP). Osteoporosis risk assessment, diagnosis and management. (RACGP, 2026).
6. Osteoporosis New Zealand. Guidance on the diagnosis and management of osteoporosis in New Zealand. 8 (New Zealand, 2017).
7. Buzasi, E., *et al.* Risk of fractures in half a million survivors of 20 cancers: a population-based matched cohort study using linked English electronic health records. *Lancet Healthy Longev* **5**, e194–e203 (2024).
8. Chang, C.J., Chan, Y.L., Pramukti, I., Ko, N.Y. & Tai, T.W. People with HIV infection had lower bone mineral density and increased fracture risk: a meta-analysis. *Arch Osteoporos* **16**, 47 (2021).
9. Frighi, V., *et al.* Incidence of fractures in people with intellectual disabilities over the life course: a retrospective matched cohort study. *EClinicalMedicine* **52**, 101656 (2022).
10. Smith, M., Roast, J., Collins, G.S., Holt, T.A. & Frighi, V. Development and external validation of prediction models for major osteoporotic fracture and hip fracture in people with intellectual disability. *Osteoporos Int* **37**, 915–925 (2026).
11. Khan, M.I. Management of bone loss due to endocrine therapy during cancer treatment. *Osteoporos Int* **34**, 671–680 (2023).
12. Ge, G., Li, J. & Wang, Q. Heart failure and fracture risk: a meta-analysis. *Osteoporos Int* **30**, 1903–1909 (2019).
13. Dong, G., Zhang, N., Wu, Z., Liu, Y. & Wang, L. Multiple sclerosis increases fracture risk: a meta-analysis. *Biomed Res Int* **2015**, 650138 (2015).
14. Hippisley-Cox, J. & Coupland, C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: Prospective open cohort study. *BMJ (Online)* **345**(2012).
15. Gregson, C.L. CFracture, an alternative to QFracture that accounts for mortality to better predict fragility fracture risk. *The Lancet Healthy Longevity* **4**, e6–e7 (2023).
16. Vyas, D.A., Eisenstein, L.G. & Jones, D.S. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. *New England Journal of Medicine* **383**, 874–882 (2020).
17. Kanis, J.A., *et al.* Race-specific FRAX models are evidence-based and support equitable care: a response to the ASBMR Task Force report on Clinical Algorithms for Fracture Risk. *Osteoporos Int* **35**, 1487–1496 (2024).
18. Burnett-Bowie, S.-A.M., *et al.* The American Society for Bone and Mineral Research Task Force on clinical algorithms for fracture risk report. *Journal of Bone and Mineral Research* **39**, 517–530 (2024).
19. Matos, J., *et al.* Critical appraisal of fairness metrics for artificial intelligence-based clinical prediction models: a scoping review. *Lancet Digit Health* **8**, 101001 (2026).
20. van der Meijden, S.L., *et al.* Navigating fairness in artificial intelligence-based prediction models: theoretical constructs and practical applications. *Lancet Digit Health*, 101035 (2026).
21. Vandenput, L., *et al.* Update of the fracture risk prediction tool FRAX: a systematic review of potential cohorts and analysis plan. *Osteoporos Int* (2022).

22. Matsushima, H., *et al.* Androgen deprivation therapy-related fracture risk in prostate cancer: an insurance claims database study in Japan. *J Bone Miner Metab* **42**, 223–232 (2024).
23. National Institute for Clinical Excellence. Prostate cancer: diagnosis and management. Vol. 2026 (London, 2019).
24. American Urological Association. Advanced Prostate Cancer: AUA/SUO Guideline (2026). Vol. 2026 (United States, 2026).
25. Suarez-Almazor, M.E., *et al.* Association of Bone Mineral Density Testing With Risk of Major Osteoporotic Fractures Among Older Men Receiving Androgen Deprivation Therapy to Treat Localized or Regional Prostate Cancer. *JAMA Netw Open* **5**, e225432 (2022).
26. Biver, E., *et al.* Diagnosis, prevention, and treatment of bone fragility in people living with HIV: a position statement from the Swiss Association against Osteoporosis. *Osteoporos Int* **30**, 1125–1135 (2019).
27. Whitney, D.G., *et al.* Elevated fracture risk for adults with neurodevelopmental disabilities. *Bone* **130**, 115080 (2020).
28. (!!! INVALID CITATION !!! 4).
29. Livingstone, S.J., *et al.* Derivation and validation of the CFrature competing risk fracture prediction tool compared with QFrature in older people and people with comorbidity: a population cohort study. *The Lancet Healthy Longevity* **4**, e43–e53 (2023).
30. Hippisley-Cox, J., Coupland, C. & Brindle, P. The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study. *BMJ Open* **4**, e005809 (2014).
31. Hippisley-Cox J. QResearch Open Code Lists. Vol. 2023 (2026).
32. Bonneville, E.F., Resche-Rigon, M., Schetelig, J., Putter, H. & de Wreede, L.C. Multiple imputation for cause-specific Cox models: Assessing methods for estimation and prediction. *Stat Methods Med Res* **31**, 1860–1880 (2022).
33. White, I.R. & Royston, P. Imputing missing covariate values for the Cox model. *Statistics in medicine* **28**, 1982–1998 (2009).
34. Rubin, D.B. *Multiple Imputation for Non-response in Surveys.*, (John Wiley, New York, 1987).
35. Hippisley-Cox J, C.C. QFrature-2016 Update information. Vol. 2026 (London, 2016).
36. Kalbfleisch, J. & Prentice, R. *The Statistical Analysis of Failure Time Data.*, (Hoboken, 2002).
37. (!!! INVALID CITATION !!! 36).
38. van Geloven, N., *et al.* Validation of prediction models in the presence of competing risks: a guide through modern methods. *BMJ* **377**, e069249 (2022).
39. Vickers, A.J., Cronin, A.M., Elkin, E.B. & Gonen, M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* **8**, 53 (2008).
40. Vickers, A.J., Van Calster, B. & Steyerberg, E.W. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* **352**(2016).
41. Graf E, Schmoor, C., Sauerbrei, W. & Schumacher, M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine* **18**, 2529–2545 (1999).
42. Gerds, T.A. & Schumacher, M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical journal. Biometrische Zeitschrift* **48**, 1029–1040 (2006).
43. Van Houwelingen, J.C. & Le Cessie, S. Predictive value of statistical models. *Statistics in medicine* **9**, 1303–1325 (1990).
44. Wolbers, M., Koller, M.T., Witteman, J.C.M. & Steyerberg, E.W. Prognostic Models With Competing Risks: Methods and Application to Coronary Risk Prediction. *Epidemiology (Cambridge, Mass.)* **20**, 555–561 510.1097/EDE.1090b1013e3181a39056 (2009).

METHODS

STUDY DESIGN AND DATA SOURCES

We undertook a population-based cohort study using two large UK primary care electronic health-record databases: QResearch and the Clinical Practice Research Datalink (CPRD Gold). QResearch comprises practices in England using the EMIS clinical computer system, whereas CPRD Gold comprises practices using Vision software. CPRD Gold practices from England and Wales provided a geographically and database-distinct external validation cohort with no overlap of GP practices or patients. Both databases are nationally representative and contain prospectively recorded information on demographics, diagnoses, prescriptions and clinical measurements³⁰. We included practices which had linked hospital admission and mortality records. We applied the same study definitions to both data sources.

Before model development, QResearch practices in England were randomly allocated at practice level to the derivation cohort (75%) or a held-out validation cohort (25%).

We included adults aged 30–100 years who were registered with a participating general practice during the study period: 1 January 2013 to 31 March 2024 for QResearch and 1 January 2001 to 31 December 2012 for CPRD to ensure a non-overlapping temporally distinct cohort.

Cohort entry was the latest of the participant's 30th birthday, one year after registration with the practice, or 1 January 2013 for QResearch and 1 January 2001 for CPRD. All participants were included in the analysis of major osteoporotic fracture. Those with prior hip fracture were excluded from the hip fracture analysis. Follow-up continued until the earliest of the relevant incident fracture outcome, death, deregistration from the practice or the database-specific study end date.

OUTCOME DEFINITIONS

The primary outcome was incident major osteoporotic fracture (MOF), defined as the fractured neck of femur (hip fracture), clinical vertebral, proximal humeral or wrist fracture after cohort entry^{14,21}. The secondary outcome was incident hip fracture. Individuals with a history of any osteoporotic fracture type before cohort entry remained eligible, with previous fracture included as a baseline predictor. Outcomes were ascertained from primary care records and linked hospital-admission and mortality datasets using published code lists based on SNOMED CT and International Classification of Diseases, 10th revision (ICD-10) codes (Supplementary Tables 10 & 11).³¹ Ascertainment of outcomes was based on pre-existing records and not affected by model predictor assessment. When the same event was recorded in more than one source, the earliest date was used, and multiple records relating to the same fracture episode were treated as one event. To account for competing risk of death we also identified non-fracture related death as the competing event.

PREDICTOR VARIABLES

Candidate predictors were prespecified and comprised established variables included in existing fracture-risk algorithms together with additional predictors selected a priori from published evidence identifying patient groups whose fracture risk is not explicitly represented in existing algorithms. Additional candidates included specific cancer types, learning disability, Down syndrome, heart failure, multiple sclerosis, visual impairment, hearing loss, HIV/AIDS, aromatase-inhibitor therapy, androgen-deprivation therapy, proton-pump inhibitor use and cystic fibrosis (Table 1, Supplementary Table 9). Ethnicity was based on self-reported ethnicity, and sex was defined as biological sex at birth as recorded in the clinical record. The objective was to estimate absolute risk of both fracture outcomes at 10 years using information available at the time of prediction, consistent with the intended clinical application of the model.

All predictors were based on information recorded in the GP record before cohort entry and assessed at cohort entry and treated as fixed baseline covariates. Medication exposure was defined as at least two prescriptions before cohort entry, including one issued within the preceding 28 days. Diagnoses were defined as ever recorded before cohort entry. Cancer predictors were based on recorded diagnoses and did not distinguish active from historical disease, metastatic status, stage or severity.

MODEL DEVELOPMENT

Missing data for ethnicity, body mass index, smoking status and alcohol consumption were imputed using multiple imputation with chained equations, incorporating all predictor variables and outcome information.^{32,33} Five imputations were generated and used in each sex/cohort analysis, reflecting the size of the datasets and available computational resources, and estimates were combined using Rubin's rules.³⁴ Binary variables for diagnoses were coded as present if a diagnosis had been recorded before cohort entry and absent otherwise.

We developed two models for each fracture outcome using separate cause-specific Cox models in men and women. Model A (the QFracture-variable competing risk model) included the predictors used in the current 2016 QFracture algorithm^{14,35}, but re-estimated their associations and incorporated competing risk of mortality. Model B (the extended model) retained these predictors and additionally considered the new candidate predictors. Candidate predictors were included in Model B if they were clinically plausible, statistically significant ($P < 0.01$) and associated with an adjusted hazard ratio below 0.90 or above 1.10, consistent with the original QFracture methodology.¹⁴ We included interactions between all predictor variables and age in Models A and B if they were statistically significant. We also fitted cause-specific Cox models for mortality (excluding deaths where fracture was specified as a cause of death) using the same predictors, fractional polynomial terms and age-interactions as for the corresponding fracture event. Estimates from the fracture and mortality models were combined to derive sex-specific equations for absolute fracture risk accounting for competing mortality.³⁶ Although the equations permit risk estimation over any period up to 10 years after cohort entry, we report 10-year risks to facilitate comparison

with existing fracture-prediction tools. The same modelling strategy was applied separately to major osteoporotic fracture and hip fracture.

MODEL EVALUATION

Multiple imputation to replace missing data was undertaken separately within each validation cohort for women and men using the same approach as for model development. Risk equations for the new models and for QFracture-2016 were applied to the held-out QResearch and external CPRD validation cohorts. Discrimination was assessed using the C-index accounting for competing risks.³⁷ Calibration was assessed using the observed cumulative incidence (estimated using the Aalen-Johansen estimator), calibration plots, ratios of observed to estimated risks and calibration intercepts and slopes based on pseudo-values which all accounted for the competing risk of non-fracture death.³⁸ Decision curve analysis accounting for competing risks compared the net benefit of Models A and B with QFracture-2016 and with “treat-all” and “treat-none” strategies.^{39,40} Model performance for Models A and B was compared with QFracture in both validation cohorts overall. Ten-year probabilistic accuracy was assessed for Model A, B and C and QFracture using inverse-probability-of-censoring-weighted Brier scores, with competing deaths treated as non-fracture outcomes.^{41,42} Subgroup analyses to assess fairness were undertaken in the QResearch held-out cohort by 10-year age bands, ethnic group and clinical conditions. Model optimism was assessed by calculating heuristic shrinkage.⁴³

As an additional analysis to examine the impact of having a model without ethnicity on the performance measures, we developed Model C which used the same modelling approach and predictors as Model B but excluded ethnicity as a predictor variable. In the held-out QResearch cohort, we compared discrimination and calibration of Model C with Model B overall and by ethnic-group separately by fracture outcome and sex, using competing-risk C-indices, observed-to-expected ratios with 95% confidence intervals, calibration intercepts and slopes, and calibration plots; consistent with recent fairness frameworks, we examined complementary probability-based performance measures rather than defining fairness using a single metric.^{19,20}

RECLASSIFICATION STATISTICS

Risk estimates at 10 years for QFracture, Model A and Model B were compared at illustrative prespecified thresholds of 10% for major osteoporotic fracture used in the National Institute of Clinical Excellence guidance in the UK²⁸ and 3% for hip fracture, commonly used in the United States³, New Zealand⁶ and Australia⁵. Clinical utility was assessed using sensitivity, specificity and positive and negative predictive values estimated using inverse probability of censoring weights at the clinical threshold. Competing deaths before 10 years were treated as known non-fracture outcomes, while independent censoring was handled using inverse-probability-of-censoring weights. This was supplemented by equal-sized reclassification analyses using risk thresholds defined to give equal numbers classified as “high risk” for each model, and comparing observed risk among participants reclassified upward and downward. These approaches follow established competing risk validation and decision-curve principles.⁴⁴

Brier scores

We compared inverse-probability-of-censoring-weighted Brier scores at 10 years between Model B and each comparator (Model A, Model C and QFracture), overall and within age groups and ethnic groups separately by sex and fracture outcome. Paired differences were calculated as the comparator-model Brier score minus the Model B Brier score, so that positive values indicated lower prediction error with Model B; 95% confidence intervals were estimated by applying the same bootstrap resamples to both models within each comparison.

All eligible individuals were used for model development and validation to maximise precision and generalisability. Analyses were performed using Stata version 18.

ETHICS

The study was approved by the East Midlands–Derby Research Ethics Committee for QResearch (OX330, reference 23/EM/0400). The project is registered here <https://www.qresearch.org/research/research-programs-and-projects/comparisons-of-risk-prediction-algorithms-using-three-clinical-research-databases-qresearch-cprd-aurum-and-cprd-gold/>. It was also approved through the CPRD research approval process (project reference 20_000162) and is registered here <https://www.cprd.com/approved-studies/evaluation-risk-prediction-algorithms-cprd-gold-and-aurum>). Individual participant consent was not required given the anonymised nature of the data.

DATA AND CODE AVAILABILITY

The authors accessed anonymised data under licence agreements with QResearch (www.qresearch.org) and CPRD (www.cprd.com). Access to the underlying patient-level data is available only through the respective data-governance processes of QResearch and CPRD. The full statistical code which implements Model A and B is published at <https://jhip.uk/X14/> under a free academic license. Clinical codes are available under a Creative Commons licence through the QResearch code-group library, <https://www.qresearch.org/data/qcode-group-library/> with accompanying details in the supplementary material. The QFracture calculator is available at the following link www.qfracture.org.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ICJMEformJHC.docx](#)
- [ICJMEformJHC.docx](#)
- [ICJMEformCACC.docx](#)
- [ICJMEformCACC.docx](#)
- [TRIPODAIChecklist.docx](#)
- [TRIPODAIChecklist.docx](#)
- [SupplementaryFiguresandTables.docx](#)
- [SupplementaryFiguresandTables.docx](#)
- [STROBEChecklist.docx](#)
- [STROBEChecklist.docx](#)