Osteoporosis International (2024) 35:1989–1998
https://doi.org/10.1007/s00198-024-07224-z

ORIGINAL ARTICLE

Fracture risk assessment in the presence of competing risk of death

Thach S. Tran[1,2,3] · Dana Bliuc[2,3] · Robert D. Blank[2] · Jacqueline R. Center[2,3,4] · Tuan V. Nguyen[1,4,5]

Received: 13 May 2024 / Accepted: 2 August 2024 / Published online: 15 August 2024

Abstract

Purpose To identify the optimal statistical approach for predicting the risk of fragility fractures in the presence of competing event of death.

Methods We used real-world data from the Dubbo Osteoporosis Epidemiology Study that has monitored 3035 elderly participants for bone health and mortality. Fragility fractures were ascertained radiologically. Mortality was confirmed by the State Registry. We considered four statistical models for predicting fracture risk: (i) conventional Cox's proportional hazard model, (ii) cause-specific model, (iii) Fine-Gray sub-distribution model, and (iv) multistate model. These models were fitted and validated in the development (60% of the original sample) and validation (40%) subsets, respectively. The model performance was assessed by discrimination and calibration analyses.

Results During a median follow-up of 11.3 years (IQR: 7.2, 16.2), 628 individuals (34.5%) in the development cohort fractured, and 630 (34.6%) died without a fracture. Neither the discrimination nor the 5-year prediction performance was significantly different among the models, though the conventional model tended to overestimate fracture risk (calibration-in-the-large index = −0.24; 95% CI: −0.43, −0.06). For 10-year risk prediction, the multistate model (calibration-in-the-large index = −0.05; 95% CI: −0.20, 0.10) outperformed the cause-specific (−0.23; −0.30, −0.08), Fine-Gray (−0.31; −0.46, −0.16), and conventional model (−0.54; −0.70, −0.39) which significantly overestimated fracture risk.

Conclusion Adjustment for competing risk of death has minimum impact on the short-term prediction of fracture. However, the multistate model yields the most accurate prediction of long-term fracture risk and should be considered for predictive research in the elderly, who are also at high mortality risk.

Summary Fracture risk assessment might be compromised by the competing event of death. This study, using real-world data found a multistate model was superior to the current competing risk methods in fracture risk assessment. A multistate model is considered an optimal statistical method for predictive research in the elderly.

Keywords Cause-specific · Competing risk · Death · Fine-Gray sub-distribution · Fracture · Multistate

Introduction

A competing risk presents when an individual is at risk of more than one mutually exclusive event and its occurrence precludes the occurrence of the primary event of interest [1, 2]. Death is a competing event in fracture risk assessment because individuals who have died would biologically have no chance of sustaining a fracture. Assessing the absolute risk of fracture is crucial in the diagnosis and management of osteoporosis as fragility fractures pose a significant public health problem worldwide [3, 4]. This assessment enables the identification of suitable preventive and therapeutic interventions that can be implemented to mitigate the risk of fracture and its associated consequences. Nevertheless, the estimation of the predicted risk of fracture becomes inaccurate in the presence of a

✉ Thach S. Tran
SonThach.Tran@uts.edu.au; th.tran@garvan.org.au

1 School of Biomedical Engineering, University of Technology Sydney, Level 10, Building 11, City Campus, Broadway, NSW 2007, Australia

2 Garvan Institute of Medical Research, Sydney, NSW, Australia

3 Faculty of Medicine, UNSW Sydney, Sydney, NSW, Australia

4 School of Medicine Sydney, University of Notre Dame Australia, Sydney, Australia

5 Tam Anh Research Institute (TAMRI), Ho Chi Minh City, Vietnam

competing risk of death as elderly individuals might die due to other reasons before experiencing a fracture [1, 2]. The adjustment for competing risk of death is therefore highly relevant for fracture assessment since fractures increase exponentially with age and the global population is aging.

At present, non-osteoporotic individuals whose 10-year risk of fracture of at least 20% are recommended for anti-fracture treatment [5, 6]. However, all existing fracture risk assessment tools have inadequate predictive capability [7]. Among the tools with the best discrimination ability, the Garvan [8] and QFracture Calculators [9] overestimate the risk of fracture in the high-risk groups, whereas the FRAX® [10] underestimates fracture risk [7]. In addition to differences in predictors, these models manage the competing death differently. While the Garvan [8] and QFracture Calculators [9] treat death without a fracture as a right-censored event, the FRAX® [10] is stated to "account for competing death" though it remains unclear how the FRAX® accounted for competing risk of death.

The most common methods for competing risk adjustment are the cause-specific hazard regression and the Fine-Gray sub-distribution hazard regression [11–13]. The cause-specific hazard regression combines both a model for fracture and that for mortality to estimate the risk of fracture [12, 13]. The Fine-Gray method assumes that individuals who have encountered a competing event are still susceptible to the primary event. This assumption, though unnatural in the context of competing death, is technically necessary to establish the one-to-one relationship with the cumulative incidence function, thus being capable of developing a model that correctly predicts the absolute risk of the primary event [14]. The multistate regression method models the progression from one state to another (e.g., from no fracture to mortality, from no fracture to fracture, or from fracture to mortality). This approach enables the competing risk of death to be naturally taken into account [15], resulting in unbiased estimates for each related outcome individually [16]. In this study, we sought to test the hypothesis that the multistate model predicts fracture risk as accurately as the most common competing risk adjustment methods. The results of this study will offer valuable insights into individualized risk assessment, thereby aiding in the identification of individuals who are at an increased risk of fracture. Using fracture risk assessment as an example, the study can inform the more general problem of predictive research in the presence of competing risks.

Methods

Study design and participants

We used data from the Dubbo Osteoporosis Epidemiology Study for which the study design and protocols have been described in detail elsewhere [17]. Briefly, through the electoral roll and via media campaign, all community-dwelling women and men aged 60 years or older as of 30 June 1989, living in Dubbo City, New South Wales, Australia, were invited to participate in the study. There is only one hospital and three radiology services for the entire Dubbo region. This centralized healthcare system, in addition to a geographically isolated research community, allows a complete ascertainment of all fractures and mortality among elderly people aged 60 years or older in the whole Dubbo region, making censoring minimal [18]. The study was approved by St. Vincent's Hospital Human Research Ethics Committee, New South Wales, Australia (HREC reference number: 13/254) and carried out according to the Australian National Health and Medical Research Council Guidelines, consistent with the Declaration of Helsinki. All participants provided written informed consent.

Regular visits were conducted biennially for a detailed and ongoing assessment of bone health. At recruitment and each visit, a nurse coordinator interviewed participants by administering a structured questionnaire to obtain anthropometric data, lifestyle factors, number of falls during the previous 12 months, prior fracture after the age of 50 years, chronic health disorders and medications prescribed. Bone mineral density (BMD) was measured at the lumbar spine and femoral neck by dual-energy x-ray absorptiometry (Lunar DPX-L; GE-Lunar).

Outcome assessments

The X-ray reports from all three radiology services for the entire Dubbo area were reviewed regularly to identify incident fractures occurring between recruitment until recently. The circumstances surrounding each fracture were determined by phone call after each fracture. The analysis included only fractures involving minimal trauma less than or equivalent to fall from standing height. High-trauma fractures, those due to underlying diseases, e.g., cancer or Paget disease, or those of digits, skull, or cervical spine were excluded. All deaths in the region were obtained from funeral lists and obituary review with verification from the State Registry of Births, Deaths, and Marriages.

Statistical analysis

As all models included five predefined predictors [8] and the study aimed to quantify the predictive performance in the validation cohort, the study population was randomly split into the development cohort (60%) and the validation cohort (40%) [19].

First, we fitted four regression models that apply different statistical methods to account for the competing death in the development cohort. They included (i) the conventional Cox's proportional hazard model, (ii) the

cause-specific hazard model, (iii) the Fine-Gray sub-distribution hazard model, and (iv) the multistate model. The conventional model estimates the risk of fracture, right censoring the competing death (i.e., the death without a fracture); whereas the cause-specific hazard, Fine-Gray and multistate models apply different methods to account for the competing death (Supplemental Methods).

Briefly, the conventional approach models fracture risk under the assumption that individuals who remain under follow-up have the same fracture risk as those who die without a fracture as if the occurrence of fracture is independent of the occurrence of death without a fracture (Figure S1A). By contrast, the cause-specific hazard approach, as the name implies, models the cause-specific hazards for fracture and those for death without a fracture separately and then combines these two models' coefficients to obtain a valid estimation of the cumulative hazard for fracture (Figure S1B) [11, 12]. The Fine-Gray method treats individuals who have died without a fracture as if they are still at risk of fracture, representing "immortal" time, but assigns a gradual reduction of weights for those with the competing death in modeling fracture risk (Figure S1C) [14]. Finally, the multistate model treats fracture and death without a fracture as two separate "states" but takes their complex inter-correlation into account (Figure S1D) [15]. Whereas the other approaches compute the cumulative incidence of fracture, the multistate model estimates the transition risk from the "event-free" state to the "fracture" state at a particular time t which is technically the fraction of individuals with a fracture at time t.

Follow-up time to fracture was calculated from the recruitment date to the date of fracture, while the follow-up time was calculated until the date of death for individuals who died without a fracture, the date of last visit or 30 June 2018, whichever came first for those who remained fracture-free. All four models used the same fracture predictors, including sex, age, femoral neck BMD, the presence of falls during the last 12 months and the presence of prior fracture after the age of 50 years prior to the study entry [8] to allow cross-comparison of their predictive performance. These predictor variables had no missing data. A proportional hazard assumption was graphically checked using the Schoenfeld residuals [20].

Secondly, we quantified the predictive accuracy of the four regression models in the validation cohort using both discrimination and calibration analyses that have been widely employed to validate the predictive accuracy of the existing fracture risk assessment tools for predicting the occurrence of fracture at clinically relevant time points [7]. Specifically, we examined the predicted absolute risks of fracture at 5 and 10 years of follow-up with the primary focus on the 10-year risk that is widely used in reality to identify high-risk individuals [5, 6, 21].

The discrimination performance was primarily quantified using Harrell's concordance C index [22] with a value closer to 1 indicating better discrimination. Harrell's C index was calculated specifically for each of the four models of interest. We used a flexible calibration curve with the addition of confidence limits for predicted group categorization [23] as the primary calibration measure for the moderate model calibration which has been shown to be realistic in epidemiologic research and considered a pragmatic guarantee that decision-making based on the model is not clinically harmful [24]. The calibration curve is constructed for centiles of predicted fracture risk with the closer concordance between the predicted fracture risks and observed fracture rates to the line of perfect prediction indicating better calibration. The predicted fracture risks were estimated from the prediction models for each participant in the validation cohort at single time point, whereas the observed fracture rates were computed as the number of participants who sustained a fracture up to the specific time point over a total of participants in each centile of predicted risk [24]. The calibration curve with its corresponding 95% confidence interval (CI) is then drawn as the average predicted fracture risk in each centile as the x-axis against the observed fracture rate in the same centile as the corresponding y-axis [24]. For quantitative comparison of the calibration performance across the models, we reported the "calibration-in-the-large" index that quantifies the overall difference between the average observed event rate and the average predictive risk [25] and the estimated calibration index that summarizes a flexible calibration curve into a single value [26]. Ideally, the calibration-in-the-large index is zero. The prediction of fracture risk was considered accurate if the average predictive values were not significantly different from the average observed fracture events (i.e., the 95% CI of the calibration-in-the-large index includes a reference unity of zero) [25]. Similarly, the estimated calibration index, calculated as the average squared difference between predicted risk and the observed event rate is zero if the flexible calibration curve is perfect. The estimated calibration index has been thus recommended as a valid measure for easily comparing calibration performance across different prediction models [26]. Other secondary measures of model's discrimination and calibration performance were also reported (Table S1).

We conducted a sensitivity analysis that mimicked the "standard" data collection in a conventional longitudinal study. The sensitivity analysis, therefore, censored the follow-up time at the last visit date and included only outcome events (i.e., fracture or death) occurring at or prior to the "hypothetically" last visit that a participant should have shown up if he or she had neither died nor been lost to follow-up (Figure S2).

The analyses were performed using the R statistical environment on a Windows platform (R-4.0.2) [27]. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Results

Baseline characteristics

The study involved 3035 participants aged 60 years or older at recruitment who were randomly split into the development and validation cohorts (Fig. 1). During a median of 11.3 years of follow-up (IQR: 7.2, 16.2), 628 (34.5%) and 630 (34.6%) participants in the development cohort fractured or died without a fracture, yielding the incidence rate of 3.35 fractures/100 person-years (95% CI: 3.10, 3.63) and 2.87 deaths/100 person-years (2.65, 3.10), respectively. Within the first 5 years of follow-up, 223 participants (12.3%) fractured and 146 (12.2%) died without a fracture.

As expected, the participants who had sustained a fracture or died without a fracture were older and had more chronic health disorders than those who remained event-free until the end of the study (Table 1). Fracture patients also had poorer bone health, more falls during the last 12 months and a greater number of fractures prior to the study entry. We found the incidence of fractures and mortality, estimated by the Kaplan–Meier method that ignores a competing risk of death was higher than those estimated by a cumulative incidence function accounting for a competing death (Figure S3).

In the development cohort, the coefficients associated with predictors of fractures were identical between the conventional Cox's model, cause-specific hazard, and multistate models (Table 2). The proportional hazard assumption was met for all four fracture risk models.

Comparison of the predictive performance between models

In the validation cohort, 418 participants (34.4%) had died and 418 participants (34.4%) had fractured within a median follow-up time of 11.1 years (IQR: 6.9, 16.6). These figures were highly comparable with the development group (Table 1).

While all prediction models had similar discrimination performance, their calibration performance for fracture prediction was affected by the prediction time (Figs. 2, 3; Table 3). Specifically, we found similar measures of discrimination performance across all prediction models, e.g., the *C*-statistic ~ 0.73 and 0.69, the Somers' D_{xy} ~ 0.47 and 0.38, and the discrimination index ~ 0.07 and 0.07 at 5- and 10-year risk assessments, respectively. By contrast, all competing risk models predicted fracture risk accurately up to 5 years, whereas the conventional Cox's model tended to overestimate the 5-year risk of fracture with the calibration-in-the-large index being − 0.24 (95% CI: − 0.43, − 0.06) (Fig. 2; Table 3). The differences in the predicted probability of fracture between the models became pronounced when

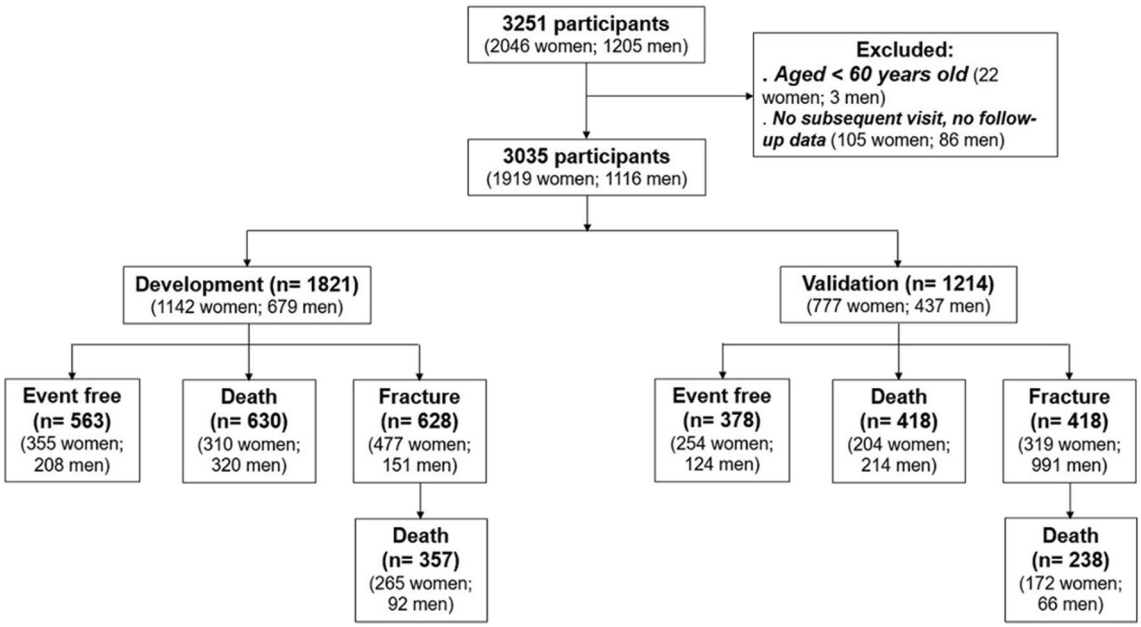

Fig. 1 Flowchart of recruitment and follow-up

Table 1 Baseline characteristics of the participants stratified by fracture and mortality status

	Development cohort ($n=1821$)			Validation cohort ($n=1214$)		
	Event-free* ($n=563$)	Fracture ($n=628$)	Death without a fracture ($n=630$)	Event-free* ($n=378$)	Fracture ($n=418$)	Death without a fracture ($n=418$)
Women	355 (63.1%)	477 (76.0%)	310 (49.2%)	254 (67.2%)	319 (76.3%)	204 (48.8%)
Age (years), mean (SD)	66.6 (4.3)	70.4 (6.9)	71.5 (6.8)	66.8 (4.8)	70.2 (6.7)	72.1 (7.3)
Femoral neck BMD T-score, mean (SD)	−0.96 (1.11)	−1.82 (1.10)	−1.29 (1.27)	−1.17 (1.09)	−1.83 (1.14)	−1.36 (1.34)
Number of falls in the last 12 months						
0	475 (84.4%)	460 (73.2%)	530 (84.1%)	314 (83.1%)	331 (79.2%)	326 (78.0%)
1	62 (11.0%)	116 (18.5%)	64 (10.2%)	42 (11.1%)	61 (14.6%)	64 (5.3%)
2 or more	26 (4.6%)	52 (8.3%)	36 (5.7%)	22 (5.8%)	26 (6.2%)	28 (6.7%)
Prior fracture after the age of 50 years prior to recruitment	55 (9.8%)	130 (20.7%)	55 (9.8%)	60 (15.9%)	78 (18.7%)	57 (13.6%)
Body mass index (kg/m^2), mean (SD)	27.5 (4.6)	26.7 (3.7)	27.5 (4.6)	27.1 (4.9)	26.5 (3.8)	26.9 (2.5)
Chronic health conditions						
Hypertension	290 (51.5%)	315 (50.2%)	308 (48.9%)	183 (48.4%)	189 (45.2%)	219 (52.4%)
Cardiovascular disease	142 (25.2%)	223 (35.5%)	261 (41.4%)	99 (26.2%)	143 (34.2%)	183 (43.8%)
Diabetes mellitus	76 (13.5%)	58 (9.2%)	88 (14.0%)	44 (11.6%)	27 (6.5%)	60 (14.4%)
Chronic respiratory disease	53 (9.4%)	79 (12.6%)	72 (11.4%)	35 (9.3%)	62 (14.8%)	38 (9.1%)
Cancer	47 (8.3%)	61 (9.7%)	51 (8.1%)	37 (9.8%)	38 (9.1%)	34 (8.1%)
Neurological disease	31 (5.5%)	45 (7.2%)	36 (5.7%)	24 (6.3%)	41 (9.8%)	25 (6.0%)
Rheumatoid arthritis	15 (2.7%)	29 (4.6%)	19 (3.0%)	7 (1.97%)	19 (4.5%)	16 (3.8%)

Data presented as number (%), unless otherwise indicated

*Participants who were alive and free of fracture until the study end

death without a fracture became prevalent (Fig. 3; Table 3). Specifically, at 10-year risk assessment, the conventional Cox's model overestimated fracture risk significantly (calibration-in-the-large index: −0.54; 95% CI: −0.70, −0.39). By contrast, the multistate model (calibration-in-the-large index: −0.05; 95% CI: −0.20, 0.10) demonstrated more accurate prediction for fracture risk than the cause-specific hazard (−0.23; −0.30, −0.08) and Fine-Gray model (−0.31; −0.46, −0.16). Similar patterns were observed from other measures of the predictive performance, such as the estimated calibration index (0.236 in the multistate model versus 1.151, 0.241, and 0.296 in the conventional, cause-specific, and Fine-Gray models, respectively) (Table 3).

The sensitivity analysis was conducted to mimic the "standard" data collection in which the outcome data of the lost-to-follow-up participants are not obtainable. As expected, the sensitivity analysis had a shorter follow-up time and fewer outcome events than the primary one. During a median follow-up of 9.7 years (IQR: 5.9, 14.5), 505 (27.7%) and 271 (14.9%) participants in the development cohort fractured or died without a fracture, respectively. Nonetheless, the outcomes of the sensitivity analysis corroborated those of the main analysis, indicating that the multistate model was capable of predicting fracture risk as accurately as the cause-specific hazard model, and more accurately than both the conventional Cox's and Fine-Gray models (Figure S4).

Discussion

The competing risk of death is an important statistical consideration in prediction research that involves elderly people who are also at high risk of death because failure to account for the competing death could result in an inaccurate estimation of the primary event risk, probably leading to treating the wrong patients. In this study, by using real-world data with a full assessment of fracture and mortality, we demonstrated that without accounting for the competing risk of death, there is a risk of overestimating the long-term probability of fracture. More importantly, we found that the multistate regression outperformed all other methods in adjusting for competing risks.

Our results confirmed a previous study which observed that an adjustment for competing risk of death, regardless of the statistical methodology, has minimum impact on the short-term (e.g., 2 years) prediction accuracy [28]. However, for long-term prediction (e.g., 10 years), failure to adjust

Table 2 Association between baseline risk factors and fracture: compare the regression coefficients between the conventional Cox's proportional hazard model with no competing risk adjustment, and the competing risk models, including the cause-specific model, the Fine-Gray model, and the multistate model

	Model with no competing risk adjustment	Model with competing risk adjustment			
	Conventional Cox's proportional hazard model	Cause-specific model		Fine-Gray model	Multistate model
Outcome of interest	Fracture	Fracture	Death	Fracture	Fracture
Predictors					
Sex (women)	1.44 (1.19, 1.74)	1.44 (1.19, 1.74)	0.58 (0.49, 0.68)	1.72 (1.45, 2.04)	1.44 (1.19, 1.74)
Age (+ 1 year)	1.03 (1.02, 1.05)	1.03 (1.02, 1.05)	1.13 (1.11, 1.14)	0.99 (0.98, 1.01)	1.03 (1.02, 1.05)
Femoral neck bone mineral density T-score (− 1)	1.50 (1.38, 1.63)	1.50 (1.38, 1.63)	0.99 (0.91, 1.07)	1.41 (1.32, 1.51)	1.50 (1.38, 1.63)
History of falls in the last 12 months	1.07 (0.94, 1.22)	1.07 (0.94, 1.23)	0.77 (0.66, 0.91)	1.20 (1.07, 1.35)	1.07 (0.94, 1.23)
Prior fracture after the age of 50 years prior to recruitment	1.92 (1.56, 2.37)	1.92 (1.56, 2.37)	1.25 (0.95, 1.63)	1.62 (1.35, 1.95)	1.92 (1.56, 2.37)
Discriminative performance					
C-statistic (95% CI)	0.70 (0.68, 0.72)	0.70 (0.68, 0.72)	0.73 (0.71, 0.75)	0.68 (0.66, 0.70)	0.71 (0.70, 0.73)

Data presented as hazard ratios (95% CI), unless otherwise indicated

for competing death resulted in an over-estimation of the predicted risk of the event of interest [1, 11–13, 29]. We also found the multistate model yielded the best calibration compared with the cause-specific hazard and Fine-Gray competing risk models in the presence of a competing event of death.

If the basic observation is considered multiple small segments of follow-up for each individual, rather than a single time-to-event or censoring point, there are alternatives to Cox's proportional hazard model. These alternatives include, among others, the Poisson regression model and the accelerated failure time model, which focus on the rate rather than the time to response. In fact, the development of FRAX was based on the Poisson regression model [10]. However, a notable advantage of the Cox model is its ability to easily produce estimates of survival probabilities in clinical studies with a well-defined common entry time for all individuals, thus utilizing a single timescale. The adjustment for competing risk of death in the Cox model is methodologically challenging. For the cause-specific hazard model, it is required that the model be fitted to not only the primary event but also to all competing risk events to obtain a valid prediction [12, 13]. Although the cause-specific coefficients are derived from the separate cause-specific models, the mathematical formula used to combine all different cause-specific hazard models for predicting the primary event is often seen as a "black box," making it difficult to communicate prediction rules [13]. Furthermore, the assumption of proportionality is not often met for all cause-specific models, and therefore, it may be necessary to recognize non-proportionality or incorporate a time-interaction term to account for the estimated effect size varying over time [12]. By contrast, the Fine-Gray model, though recommended for predictive research in the presence of a competing death, relies on a counterintuitive assumption that individuals who have died are still at risk of the primary event [11, 12, 14]. Furthermore, it is not possible to directly link the effect of the risk factors for fracture on the underlying fracture risk in real-world scenarios [30]. As argued by Fine and Gray [14], the individuals who have experienced the competing risk event might be viewed as a "placeholder" for a fraction of the population who cannot experience the primary event and, as such, can constrain the definition of the sub-distribution hazard function. To account for the competing risk, the sub-distribution hazard for the primary event (i.e., fracture) among individuals who have experienced the competing event is calculated with a gradual reduction in weight over time [14].

In view of the above conceptual and methodological challenges, the multistate model appears to be superior to the cause-specific hazard and Fine-Gray models in fracture risk assessment for the elderly who are also at high risk of death. It is widely recognized that patients with a history of

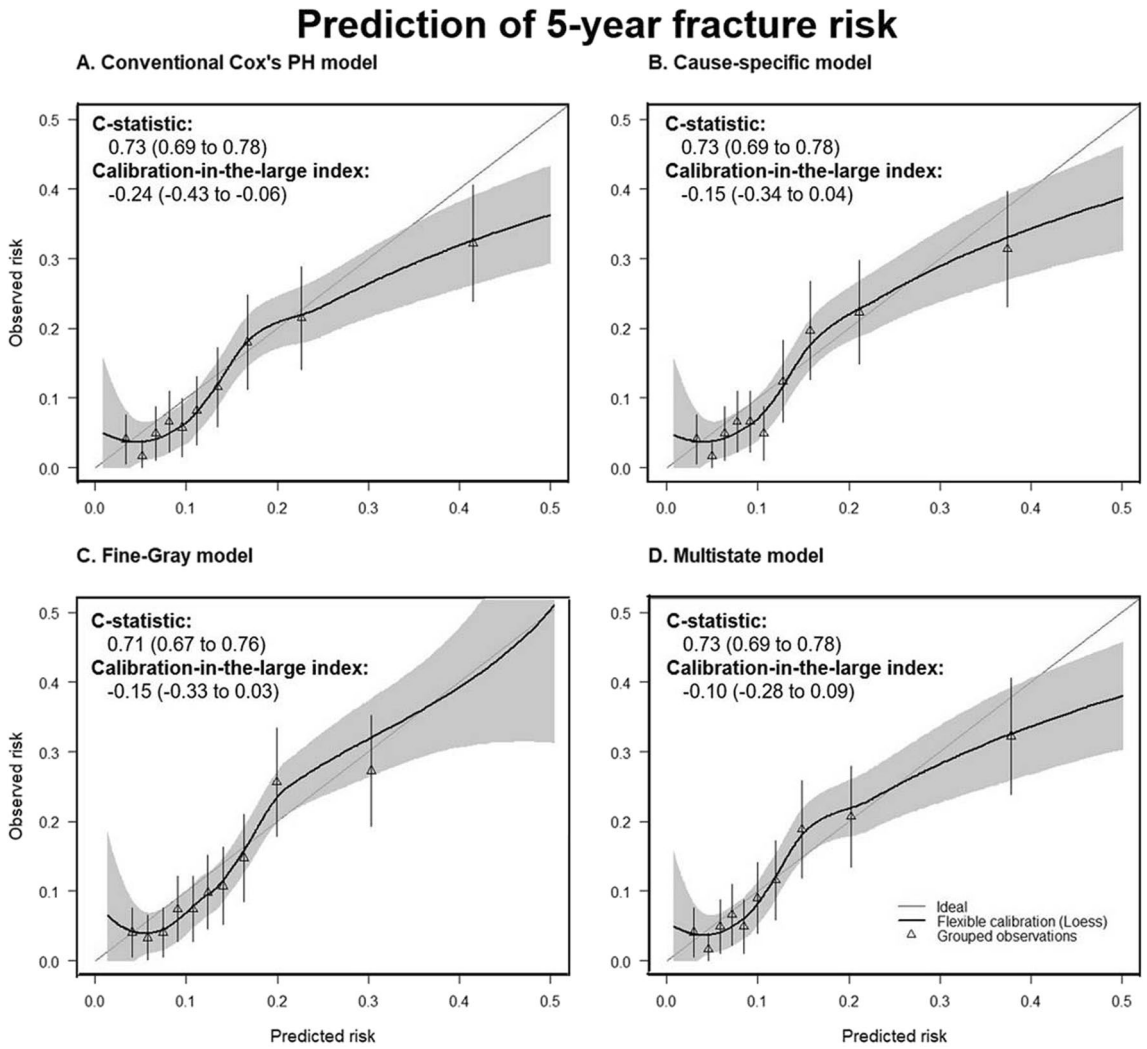

Fig. 2 Prediction accuracy of the conventional Cox's proportional hazard model with no competing risk adjustment, and the competing risk models, including the cause-specific model, the Fine-Gray model, and the multistate model: 5-year fracture risk

fractures are at greater risk of mortality compared to those without a fracture. Furthermore, those with a fracture have a higher likelihood of sustaining a subsequent fracture, which significantly increases their risk of mortality [3]. The multistate model enables the estimation of the probability of transitioning from one state to another, which is highly relevant in osteoporosis research [31, 32]. It is also easy to formulate a multistate model to estimate predicted risk based on a specific risk profile, making risk prediction rules more easily communicated than with the cause-specific hazard model. The assumption of the multistate model is more intuitive than the Fine-Gray model as the population at risk in the multistate model does not include individuals who have died. Most importantly, the multistate model is capable of estimating the risk of other correlated events within a single framework, providing adequate consideration for the complex intercorrelations among the events of interest, including competing risks [16]. For instance, the multistate model can predict not only the risk of fracture, or death without a fracture but also the consequences of fracture such as a subsequent fracture and post-fracture mortality [31, 32]. Therefore, we consider that the multistate model is a method of choice for the assessment of fracture risk in the elderly who are also at high risk of mortality.

Using fracture as an example, our study shed light on the more general issue of prediction of an adverse health event in the presence of competing risk of death in predictive research on aging. The multistate model should be considered one of preferred statistical approaches and all competing risk models examined in this study should be considered and compared for predictive research in the presence of a competing risk. Specifically, the findings can be used to extend the current recommendation [33] that the most rigorous scientific approach to analyzing competing

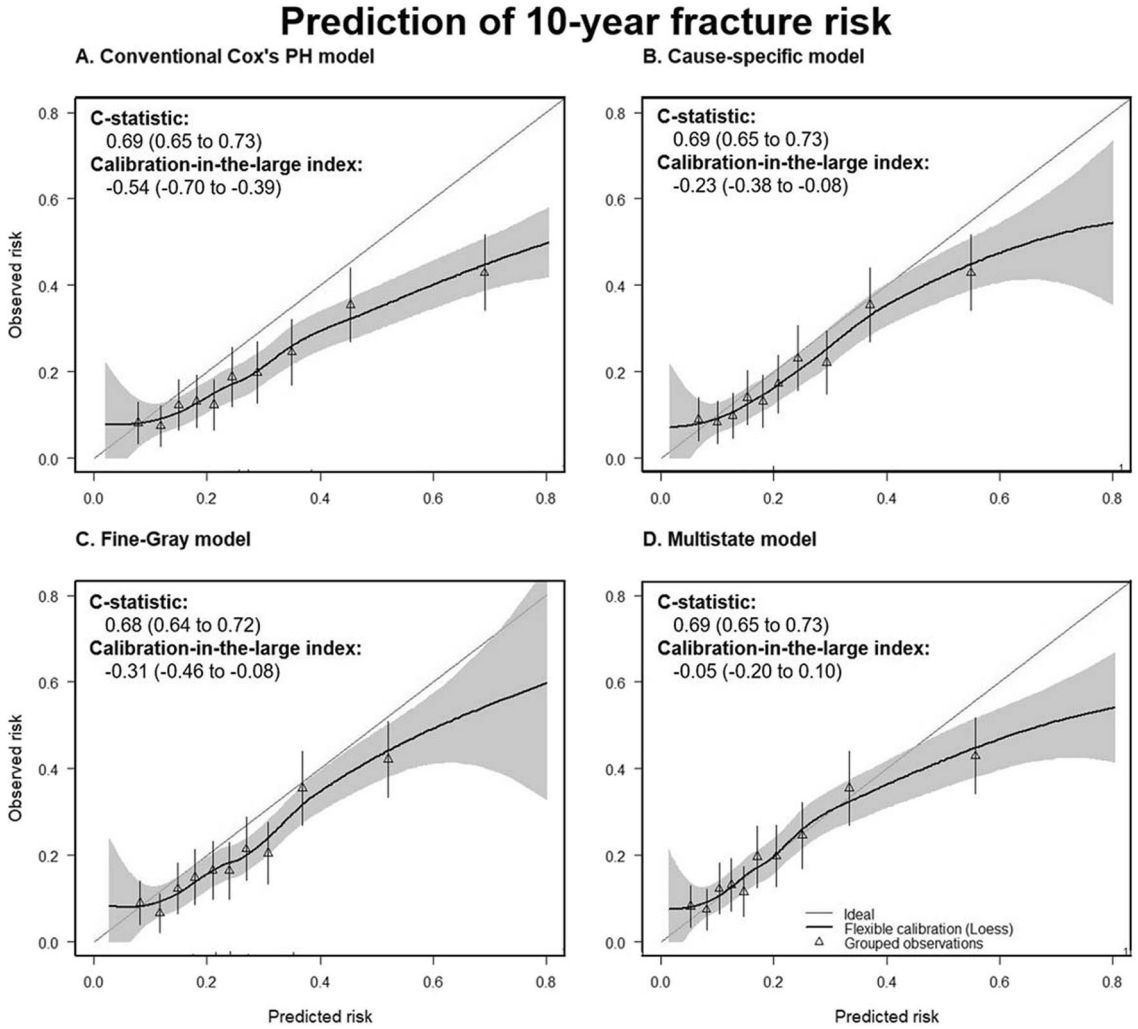

Fig. 3 Prediction accuracy of the conventional Cox's proportional hazard model with no competing risk adjustment, and the competing risk models, including the cause-specific model, the Fine-Gray model, and the multistate model: 10-year fracture risk

risk data should require not only the cause-specific hazards from the cause-specific model, the cumulative incidence function from the Fine-Gray model but also the transition risks from the multistate model be analyzed and reported side by side. The results from different competing risk models are expected to complement each other and provide more thorough understanding of the event dynamics.

Our findings should be considered within the context of their strengths and limitations. This is the first methodology study that compared the predictive performance of the multistate model against the current competing risk models. The analysis was conducted using real-world data of more than 3000 elderly people whose health status has been monitored for a median of 11 years, providing ample time for the majority of participants to experience either the primary event of fracture or the competing event of death. Another strength of this study is that it is conducted in a geographically isolated setting with a centralized healthcare system, enabling a comprehensive assessment of outcome events and capturing data for participants who did not present for subsequent visits and making censoring minimal. Additionally, the sensitivity analysis mimicking conventional cohort studies yielded consistent findings with the primary analysis, further validating the robustness of our results.

A potential limitation of the study is its limited generalizability since the study population was primarily Caucasian and from an industrialized country with a low mortality rate. However, as the predictive performance was made between different models in the same study context, i.e., the same predictors and the same study population, it is unlikely that the overall trends observed in our findings would differ significantly in other settings with different ethnicities and/or higher mortality rates.

Table 3 Performance measures of the conventional Cox's proportional hazard model with no competing risk adjustment, and the competing risk models, including the cause-specific model, the Fine-Gray model, and the multistate model

	5-year prediction of fractures				10-year prediction of fractures			
	Conventional	Cause-specific	Fine-Gray	Multistate	Conventional	Cause-specific	Fine-Gray	Multistate
Discriminative performance								
Harrell's *C*-index	0.734	0.734	0.713	0.734	0.692	0.692	0.681	0.692
R^2	0.125	0.127	0.104	0.126	0.108	0.113	0.101	0.110
Somers' D_{xy}	0.467	0.468	0.426	0.468	0.384	0.385	0.363	0.384
Discrimination index	0.065	0.066	0.054	0.065	0.069	0.073	0.065	0.071
Calibration performance								
Calibration intercept	−0.243	−0.150	−0.153	−0.098	−0.542	−0.226	−0.311	−0.050
Estimated calibration index (ECI)	0.222	0.108	0.065	0.136	1.151	0.241	0.296	0.236
Calibration slope	0.879	0.952	1.128	0.897	0.647	0.840	0.919	0.723
Brier's score	0.095	0.094	0.096	0.095	0.156	0.147	0.149	0.147
Unreliability index	0.005	0.0006	0.001	0.00004	0.057	0.008	0.014	0.008
E_{max}	0.124	0.061	0.037	0.081	0.280	0.122	0.106	0.154
E_{avg}	0.028	0.022	0.023	0.022	0.083	0.037	0.049	0.019

Harrell's *C*-index (~discriminative performance to differentiate individuals with and without the outcome), R^2 (~improvement of the model versus the null model), Somers' D_{xy} (~the rank difference between all predicted probabilities and all observed outcomes): the higher value indicates better discrimination

Discrimination index (~difference in quality between the best constant predictor and the best-calibrated predictor): the closer value to zero indicates better discrimination

Calibration intercept (~absolute difference between the average rate of observed outcomes and the average predicted risks), estimated calibration index (~average squared difference between predicted risk and observed risk), Brier's core (~mean squared error between predicted risk and observed outcome), unreliability index (~how far the logit calibration intercept and slope are from (0,1)), E_{max} (~maximum absolute difference in raw predicted and calibrated probabilities), and E_{avg} (~the average difference in raw predicted and calibrated probabilities): the closer value to zero indicates better calibration

Calibration slope (~spread of the predicted risks): the closer value to one indicated better calibration

In conclusion, our findings suggest that for long-term assessment, the adjustment for competing risk of death is necessary to produce an unbiased and accurate predicted fracture risk, and the multistate model should be considered one of the preferred statistical methods for adjusting competing risk in predictive research in the elderly.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00198-024-07224-z.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. The Dubbo Osteoporosis Epidemiology Study was supported by the National Health Medical Research Council Australia Projects 1070187, 1008219, and 1073430. Other funding bodies were an Osteoporosis Australia-Amgen grant and the Mrs. Gibson and Ernst Heine Family Foundation. Dr Tuan V Nguyen is supported by a grant from the National Health and Medical Research Council in Australia (APP1195305).

Data availability The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request. The statistical analysis plan, the sample of datasets (~10% of the actual dataset) and the R codes utilized to conduct the data analysis and generate the corresponding analysis output are publicly accessible (https://github.com/ThachSTran/Multistate-model-and-competing-death.git).

Declarations

Conflicts of interest Thach Tran and Dana Bliuc have no competing interests to declare. Robert D. Blank was an advisory member for Amgen, has consulted for Bristol Myers Squibb and Guidepoint Advisors, has received editorial stipend from Elsevier and the Endocrine Society, has had ownership in Abbott Labs, Abbvie, Amgen, Doctorpedia, GlaxoSmithKline, JangoBio, Johnson & Johnson, NovoNordisk, Pfizer, Procter & Gamble and ROMTech, and royalties from Wolters Kluwer. Jacqueline R. Center has consulted for and/or given educational talks for Amgen, Actavis, and Bayer. Tuan V. Nguyen has received honoraria for consulting and symposia from Merck Sharp and Dohme, Roche, Servier, Sanofi-Aventis, and Novartis.

References

1. Berry SD, Ngo L, Samelson EJ et al (2010) Competing risk of death: an important consideration in studies of older adults. J Am Geriatr Soc 58(4):783–787. https://doi.org/10.1111/j.1532-5415.2010.02767.x
2. Koller MT, Raatz H, Steyerberg EW et al (2012) Competing risks and the clinical community: irrelevance or ignorance? Stat Med 31(11–12):1089–1097. https://doi.org/10.1002/sim.4384
3. Bliuc D, Nguyen ND, Milch VE, Nguyen TV, Eisman JA, Center JR (2009) Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women. JAMA 301(5):513–521. https://doi.org/10.1001/jama.2009.50
4. Budhia S, Mikyas Y, Tang M et al (2012) Osteoporotic fractures: a systematic review of U.S. healthcare costs and resource utilization. PharmacoEconomics 30(2):147–70. https://doi.org/10.2165/11596880-000000000-00000
5. Camacho PM, Petak SM, Binkley N et al (2020) American Association of Clinical Endocrinologists/ American College of Endocrinology Clinical Practice Guidelines for the Diagnosis and Treatment of Postmenopausal Osteoporosis-2020 Update. Endocr Pract 26(Suppl 1):1–46. https://doi.org/10.4158/GL-2020-0524SUPPL
6. The Royal Australia College of General Practitioners (2010) Clinical guideline for the prevention and treatment of osteoporosis in postmenopausal women and older men. RACGP, South Melbourne, Vic
7. Beaudoin C, Moore L, Gagne M et al (2019) Performance of predictive tools to identify individuals at risk of non-traumatic fracture: a systematic review, meta-analysis, and meta-regression. Osteoporos Int 30(4):721–740. https://doi.org/10.1007/s00198-019-04919-6
8. Nguyen ND, Frost SA, Center JR et al (2008) Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. Osteoporos Int 19(10):1431–1444. https://doi.org/10.1007/s00198-008-0588-0
9. Hippisley-Cox J, Coupland C (2012) 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 344:e3427. https://doi.org/10.1136/bmj.e3427
10. Kanis JA, Oden A, Johnell O et al (2007) The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. Osteoporos Int 18(8):1033–1046. https://doi.org/10.1007/s00198-007-0343-y
11. Austin PC, Lee DS, Fine JP (2016) Introduction to the analysis of survival data in the presence of competing risks. Circ 133(6):601–609. https://doi.org/10.1161/CIRCULATIONAHA.115.017719
12. Lau B, Cole SR, Gange SJ (2009) Competing risk regression models for epidemiologic data. Am J Epidemiol 170(2):244–256. https://doi.org/10.1093/aje/kwp107
13. Wolbers M, Koller MT, Witteman JC et al (2009) Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 20(4):555–561. https://doi.org/10.1097/EDE.0b013e3181a39056
14. Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 94(446):496–509
15. Andersen PK, Keiding N (2002) Multi-state models for event history analysis. Stat Methods Med Res 11(2):91–115. https://doi.org/10.1191/0962280202SM276ra
16. Meira-Machado L, de Una-Alvarez J, Cadarso-Suarez C et al (2009) Multi-state models for the analysis of time-to-event data. Stat Methods Med Res 18(2):195–222. https://doi.org/10.1177/0962280208092301
17. Nguyen T, Sambrook P, Kelly P et al (1993) Prediction of osteoporotic fractures by postural instability and bone density. BMJ 307(6912):1111–1115. https://doi.org/10.1136/bmj.307.6912.1111
18. Center JR (2017) Fracture burden: what two and a half decades of Dubbo Osteoporosis Epidemiology Study data reveal about clinical outcomes of osteoporosis. Curr Osteoporos Rep 15(2):88–95. https://doi.org/10.1007/s11914-017-0352-5
19. Joseph R (2022) Optimal ratio for data splitting. Stat Anal Data Min 15(4):531–538. https://doi.org/10.1002/sam.11583
20. Schoenfeld D (1982) Partial residuals for the proportional hazards regression models. Biometrika 69(1):239–241. https://doi.org/10.1093/biomet/69.1.239
21 National Cholesterol Education Program (NCEP) (2002) Expert panel on detection, and treatment of high blood cholesterol in adults (adult treatment panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III) final report. Circulation 106(25):3143–421. https://doi.org/10.1161/circ.106.25.3143
22. Harrell FE, Califf RM, Pryor DB et al (1982) Evaluating the yield of medical tests. JAMA 247(18):2543–2546. https://doi.org/10.1001/jama.1982.03320430047030
23. Austin PC, Steyerberg EW (2014) Bootstrap confidence intervals for loess-based calibration curves. Stat Med 33(15):2699–2700. https://doi.org/10.1002/sim.6167
24. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW (2016) A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 74:167–176. https://doi.org/10.1016/j.jclinepi.2015.12.005
25. Steyerberg EW, Vickers AJ, Cook NR et al (2010) Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21(1):128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2
26. Van Hoorde K, Van Huffel S, Timmerman D, Bourne T, Van Calster B (2015) A spline-based tool to assess and visualize the calibration of multiclass risk predictions. J Biomed Inform 54:283–293. https://doi.org/10.1016/j.jbi.2014.12.016
27. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020 URL https://www.R-project.org
28. Al-Wahsh H, Tangri N, Quinn R et al (2021) Accounting for the competing risk of death to predict kidney failure in adults with stage 4 chronic kidney disease. JAMA Netw Open 4(5):e219225. https://doi.org/10.1001/jamanetworkopen.2021.9225
29. Putter H, Fiocco M, Geskus RB (2007) Tutorial in biostatistics: competing risks and multi-state models. Stat Med 26(11):2389–2430. https://doi.org/10.1002/sim.2712
30. Andersen PK, Keiding N (2012) Interpretability and importance of functionals in competing risks and multistate models. Stat Med 31(11–12):1074–1088. https://doi.org/10.1002/sim.4385
31. Tran T, Bliuc D, Pham HM et al (2020) A risk assessment tool for predicting fragility fractures and mortality in the elderly. J Bone Miner Res 35(10):1923–1934. https://doi.org/10.1002/jbmr.4100
32. Ho-Le TP, Tran TS, Bliuc D et al (2021) Epidemiological transition to mortality and refracture following an initial fracture. Elife 10:661142. https://doi.org/10.7554/eLife.61142
33. Latouche A, Allignol A, Beyersmann J et al (2013) A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. J Clin Epidemiol 66(6):648–653. https://doi.org/10.1016/j.jclinepi.2012.09.017