Age and Ageing 2025; **54:** afaf077
https://doi.org/10.1093/ageing/afaf077

RESEARCH PAPER

Development and external validation of the electronic frailty index 2 using routine primary care electronic health record data

Kate Best[1], Farag Shuweihdi[2], Juan Carlos Bazo Alvarez[3], Samuel Relton[2], Christina Avgerinou[4], Danielle Nimmons[5], Irene Petersen[3,6], Maria Pujades-Rodriguez[7], Simon Paul Conroy[8], Kate Walters[3], Robert M. West[9], Andrew Clegg[10]

[1]University of Leeds, Leeds Institute of Health Sciences - Academic Unit for Ageing and Stroke Research, Leeds, West Yorkshire, United Kingdom
[2]University of Leeds, Leeds Institute of Health Sciences, Leeds, West Yorkshire, United Kingdom
[3]University College London, Research Department of Primary Care and Population Health, London, United Kingdom of Great Britain and Northern Ireland
[4]University College London, Department of Primary Care and Population Health, London, United Kingdom
[5]University College London Faculty of Population Health Sciences, Research Department of Primary Care and Population Health, London, United Kingdom
[6]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark
[7]University of Leeds, Leeds Institute of Health Sciences, Leeds, United Kingdom
[8]The Royal London Hospital, Barts Health NHS Trust, London, United Kingdom
[9]University of Leeds, Leeds Institute of Health Sciences – Biostatistics, Leeds, West Yorkshire, United Kingdom
[10]University of Leeds - Academic Unit for Ageing & Stroke Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, West Yorkshire, United Kingdom

Address correspondence to: Andrew Clegg, University of Leeds - Academic Unit for Ageing & Stroke Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, BD9 6RJ, West Yorkshire, United Kingdom. Email: a.p.clegg@leeds.ac.uk

Abstract

Background: The electronic frailty index (eFI) is nationally implemented into UK primary care electronic health record systems to support routine identification of frailty. The original eFI has some limitations such as equal weighting of deficit variables, lack of time constraints on variables known to resolve and definition of frailty category cut-points. We have developed and externally validated the eFI2 prediction model to predict the composite risk of home care package; hospital admission for fall/fracture; care home admission; or mortality within one year, addressing the limitations of the original eFI.
Methods: Linked primary, secondary and social care data from two independent retrospective cohorts of adults aged ≥65 in 2018 was used; the population of Bradford using the Connected Bradford dataset (development cohort, 78 760 patients) and the population of Wales, from the Secure Anonymised Information Linkage databank (external validation cohort, 660 417 patients). Candidate predictors included the original eFI variables, supplemented with variables informed by literature reviews and clinical expertise. The composite outcome was modelled using Cox regression.
Results: In internal validation the model had excellent discrimination (C-index = 0.803, Nagelkerke's $R^2 = 0.0971$) with good calibration (Calibration slope = 1.00). In external validation, the model had good discrimination (C-index = 0.723, Nagelkerke's $R^2 = 0.064$), with some evidence of miscalibration (Calibration slope = 1.104).
Conclusions: The eFI2 demonstrates robust prediction for key frailty-related outcomes, improving on the original eFI. Our use of novel methodology to develop and validate the eFI2 will advance the field of frailty-related research internationally, setting a new methodological standard.

Keywords: frailty; frailty index; electronic health records; older people

Key Points

- The original electronic frailty index (eFI) has been extensively validated but has some important limitations such as equal weighting of deficit variables
- The electronic frailty index version 2 (eFI2) weights deficit variables, includes time constraints and frailty category cut-points mapped from a reference standard.
- eFI2 has robust performance for predicting home care, hospital admission for fall/fracture, care home admission, or mortality.
- eFI2 has better overall performance than the original eFI.

Background

Maximising independence through proactive prevention, identification and management of frailty is identified as a key priority in the 2023 UK Chief Medical Officer's Report on Health in an Ageing Society [1]. Frailty negatively impacts on quality of life, caregiver burden and health and social care use [2]. It accounts for £6 billion of annual NHS expenditure and is the strongest predictor of social care costs [3]. There is also evidence that outcomes for older people with frailty may be improved following community-based interventions such as comprehensive geriatric assessment and resistance exercise training, increasing the importance of identifying and targeting those with frailty in primary care [4–7].

UK and international guidelines support routine frailty identification for timely and targeted proactive care [8–10]. There is growing support for routine frailty identification to enable holistic, person-centred care in perioperative, cardiovascular, and oncology care settings [11–13]. To support routine frailty identification, the electronic frailty index (eFI) was nationally implemented into UK primary care electronic health record (EHR) systems in 2016 [14]. This was a global first, supporting national health policy change through the 2017/18 primary care contract [15], with subsequent inclusion in national medicines optimisation policy [16].

The eFI was developed using the cumulative deficit model, where the eFI score is calculated as an equally weighted proportion of deficits experienced by a patient out of the total possible [2]. The cumulative deficit model provides simple, intuitive frailty indices, but the prognostic performance may be limited by the equal weighting of deficit variables [17]. Although the discrimination of the eFI was good for 1-year mortality (C-index = 0.72 and 0.76 in internal and external validation, respectively), care home admission (C-index = 0.74 in internal validation) and emergency hospitalisation (C-index = 0.66 and 0.71 in internal and external validation) [14], it was not validated for key frailty outcomes such as loss of independence or falls [14], and constraints were not applied to variables that could resolve over time. Finally, the original eFI frailty categories were developed using a pragmatic statistical approach, splitting the space between zero and the 99th centile eFI score into categories (fit, mild frailty, moderate frailty, severe frailty), meaning that the generated frailty categories did not necessarily align with reference standard measures.

The aim of this study was to develop and externally validate the eFI version 2 (eFI2), as a clinical prediction model to predict key frailty-related outcomes that can be automatically populated from routine primary care EHR data.

Methods

Sources of data

A retrospective cohort study was used to develop the eFI2 clinical prediction model using Connected Bradford, which includes linked primary, secondary and social care data from ~800 000 residents of Bradford District and Craven, UK [18]. External validation was in a second retrospective cohort using the Secure Anonymised Information Linkage (SAIL) databank, which contains ~5 000 000 anonymised records from residents of Wales, with linked primary care, ED attendance, hospital admissions, care homes and Office for National Statistics (ONS) mortality data [19].

Participants

Patients aged ≥65 contributing data to Connected Bradford or SAIL on 1 April 2018 were included. A lookback period included the complete primary care EHR from first registration. People in receipt of a home care package and care home residents prior to 1 April 2018 were excluded from the development cohort, given that they would be unable to experience key outcomes of interest (new home care package, care home admission) and may therefore deflate the risk of the composite outcome if included. External validation is intended to assess the validity of a model in practice; therefore our external validation was performed including patients with existing home care package or who were already resident in a care home.

Outcome

The binary outcome was any one (or more) of the following four outcomes occurring within 12 months of cohort entry (1 April 2018):

1. Hospitalisation with fall or fragility fracture (as indicators of a serious fall), defined using established code lists.

2. Care home admission, identified in Connected Bradford using a care home residence flag in linked local authority data, or using the linked care homes dataset in SAIL.
3. All-cause mortality, identified using date of death in Connected Bradford and linked ONS data in SAIL.
4. New home care package, identified in Connected Bradford using a home care receipt flag in the linked Local Authority data. Home care package is not recorded in SAIL and so this was not included in the composite outcome for the external validation.

The four outcomes were selected with input from a Patient and Public Involvement (PPI) Frailty Oversight Group (FOG), because they best represent frailty-related adverse outcomes and loss of independence.

Predictors

Candidate predictors included the 36 deficit variables from the original eFI in addition to those identified by a systematic review of prognostic factors in older people with frailty [20], targeted scoping reviews of the literature, and consultation with clinical practitioners. We restricted our candidate predictors to those that are routinely recorded in primary care. Systematised Nomenclature of Medicine Clinical Terms (SNOMED CT) code lists for each candidate predictor were developed from the original eFI, National Health Service (NHS) Quality and Outcomes Framework incentivisation scheme code lists (2018/19), [21] and from established code lists. Mapping between coding ontologies (SNOMED CT used in Connected Bradford and Read version 2 in SAIL) was via NHS Technology Reference Update Distribution mapping tables. Where mapping was unsuccessful, manual searching for an appropriate Read code was conducted via searching the code's text description in www.athena.ohdsi.org. There were a small number of SNOMED CT codes that could not be mapped to an appropriate READ code, and these codes were therefore not included. Mapped codes were checked by clinical authors A.C., K.W., S.C., D.N. and C.A.

There were 79 candidate predictor variables (Supplementary Materials, Appendix 1). Of the 79 candidate predictors, 70 were binary and four had multiple ordered categories (alcohol intake (reference category = zero intake), body mass index (BMI, reference category = Recommended BMI), smoking status (reference category = none, ex, or missing), and polypharmacy (reference category 0–4 medications). All predictors were derived based on the presence of a relevant SNOMED CT or READ code.

Several predictors had additional inclusion rules, such as occurrence only in the previous 5 years (Supplementary Materials, Appendix 1). We excluded candidate predictors with low prevalence in the development cohort (<0.05%). Age was not included as a predictor because frailty is considered to identify increased risk of adverse outcomes for people compared to same-age peers, aligning with well-established theoretical understanding of frailty.

Sample size

Using Riley et al's 'pmsampsi' package in Stata, a prespecified sample size calculation for model development was derived [22]. Based on existing estimates of one year mortality, hospitalisation for falls and care home admission (4%, 5% and 0.7%, respectively [14, 23]) we estimated between 6% and 10% of patients would experience our composite outcome. Conservatively using the lower range of the estimated incidence, assuming a maximum of 90 predictor parameters would be included in the final model, with a Cox-Snell R-squared of at least 5%, 15 746 patients were required for model development.

Prediction model sample size guidance recommends tailoring the sample size required for external validation based on the model being validated [24]. We adapted Riley et al's Stata script for time-to-event models [24] to estimate that a sample size of 60 000 patients (with at least 4016 experiencing the composite outcome), is sufficient to target a 95% confidence interval of width 0.2 around the estimate of the calibration slope. The available sample sizes in both cohorts exceeded the sample size estimates required.

Missing data

Decisions on missing data were taken in collaboration with wider stakeholders, including suppliers of EHR systems, so that the final eFI2 model would be suitable for implementation and rollout. For binary predictors, patients that did not have a relevant code recorded in their primary care EHRs were assumed not to have the corresponding condition. Similarly, only recorded medications contributed to the polypharmacy variable. Missing lifestyle data for BMI, smoking status and alcohol intake were represented by a 'missing' category so that when the eFI2 is implemented in primary care EHR systems, patients who have missing lifestyle data in practice will still have an eFI2 score generated.

Statistical analysis

Time until the composite outcome was modelled using a multivariable Cox proportional hazards model with all predictors included. Using the penalised package in R [25], we applied a positivity constraint to shrink predictor parameters with negative coefficients (i.e. apparently 'protective' variables) to zero, in line with established definition of a frailty index deficit variable [26].

Each regression coefficient was rescaled, dividing the coefficient by the sum of coefficients. For categorical predictors with more than two categories, the category with the largest coefficient contributed to the sum of coefficients. The eFI2 score was calculated for each patient as the sum of the total possible rescaled coefficients. The rescaling of the coefficients to a 0–1 scale makes the interpretation of the eFI2 score comparable to the original eFI score, and the conventional method of presenting an FI score.

Discrimination (C-index), Nagelkerke R^2, the Expected/Observed (E/O) ratio and the calibration slope were

calculated in the development and external validation cohorts, with these being estimated using bootstrapping (50 bootstraps) in the internal validation. Further information on the interpretation of C-index, Calibration slope and E/O can be accessed elsewhere [27].

In the development and external validation cohorts, calibration performance was examined using calibration plots that compared the observed and predicted risk of the composite outcome at one year, with the predicted risk being the exponentiated linear predictor multiplied by the baseline hazard derived from the development cohort data [28]. Calibration plots included a smoothed calibration curve, estimated using a cubic spline smoother.

Selection of thresholds for frailty risk groups

Cut-points for the original eFI were calculated by splitting the distance between 0 and the 99th centile eFI score in the development cohort into four equally spaced categories. The original eFI cut-point distinguishing between fit and mild frailty was often clinically scrutinised by clinicians for not matching the clinical view of mild frailty. Therefore, the cut-points between the 'robust' and 'mild frailty' categories for the eFI2 score were assigned based on mapping from the cut-point of a research standard frailty index, using data from the Community Ageing Research 75+ (CARE75+) national longitudinal cohort study [29]. The eFI2 scores could be calculated for 267 CARE75+ participants with linked primary care electronic health record data across five time points (1316 observations). The reference standard FI score was derived using the cumulative deficit model with 60 deficits, with a score ≥ 0.25 representing any degree of frailty [30]. Therefore, we mapped this FI value to the corresponding eFI2 value using linear regression. The eFI2 cut-points that distinguish between mild, moderate and severe frailty were assigned in the external validation cohort by dividing the space between the mapped mild frailty cut-point and the 99th centile eFI2 score into three.

Comparison with the eFI

The original eFI score was calculated in the external validation cohort and model performance measures (C-index, Nagelkerke's R^2) estimated using the composite outcome.

All analyses were performed in R (R Core Team 2020).

Patient, public and practitioner involvement

Our PPI FOG, established as part of the National Institute for Health and Care Research Applied Research Collaboration Yorkshire & Humber (NIHR ARC YH) [31], helped us develop our research questions and informed the delivery of the project, including the selection of outcomes that best represent frailty. The list of potential deficits for the eFI2 was reviewed by the Frailty Clinical Network-London Region, a multi-disciplinary group representing Geriatric Medicine, General Practice, Community and Acute Nursing, Mental Health, Ambulance Service, Occupational Therapy, Community Physiotherapy, Pharmacy, Third Sector, Patient Public Voice and Social Care. The group reviewed the content of the eFI and proposed additional variables to include, highlighting consideration of chronic pain and mental health variables. Both the Frailty Clinical Network and FOG advocated weighting variables in eFI2, and identified which variables they felt could potentially resolve within a 5-year time horizon.

Results

Study population and characteristics

There was a lower one-year incidence of the composite outcome in the development cohort compared to the external validation (6.1% vs 8.5%), which was partially due to the exclusion of 2925 (3.5%) patients with previous home care or care home admission in the development cohort, and partially due to a lower incidence of hospital admissions for falls and fragility fractures in the development cohort (2.6% vs 4.9%). Cohort demographics are shown in Table 1. The prevalence of each predictor in the development and external validation cohorts is shown in Table 2.

Model development

Thirty-six predictors were included in the final eFI2 model, with 42 of the original 79 candidate predictors removed by the positivity constraint, and one candidate predictor (managing finances) excluded due to low prevalence in the development cohort (<0.05%). The median predicted risk was 3.0%, with the minimum predicted risk being 1.51% (the baseline hazard) and the 99th centile being 44%. The hypothetical total possible sum of coefficients was 8.429, meaning the transformed coefficients were calculated as the coefficient divided by 8.429 (Table 3). The median eFI2 score, calculated using the transformed coefficients, was 0.083 (IQR = 0.042–0.16) in the development cohort.

Internal validation using development cohort

The model showed excellent discrimination (boot-strapped C-index = 0.803), with 9.7% of the variation being explained by the predictors (bootstrapped Nagelkerke's $R^2 = 0.0971$) (Table 4). The average predicted risk of the composite outcome was 2.3% lower than observed risk (bootstrapped E/O = 0.987), and while the calibration slope was 1.000, there was slight underestimation where the predicted risk was between 10% and 30%, and some slight over-estimation of risk where the predicted risk was > 30% (Figure 1a). Calibration plots by age and sex are presented in Supplementary Materials, Appendix 5.

External validation

The model had good discrimination in the external validation (C-index = 0.723, 95% CI: 0.721, 0.725) (Table 4). The model explained 6.4% of the variation in the time until the composite outcome (Nagelkerke's $R^2 = 0.064$). However, on average the model tended to underestimate the risk of the composite outcome (E/0 = 0.443). There was some

Table 1. Baseline characteristics and outcome incidence in development and external validation cohorts.

Characteristic/outcome	Development cohort *N* = 78 760	External validation cohort *N* = 660 417
Age (years), mean (SD)	75.1 (7.5)	74.9 (7.5)
Gender, *n* (%)		
Male	36 345 (46.1)	311 742 (47.2)
Female	42 415 (53.9)	348 675 (52.8)
IMD quintile*, *n* (%)		
1 (most deprived)	17 751 (28.9)	85 421 (16.1)
2	11 618 (18.9)	102 410 (19.3)
3	13 806 (22.5)	108 496 (20.4)
4	10 896 (17.7)	111 726 (21.1)
5 (least deprived)	7401 (12)	122 619 (23.1)
Missing	17 288 (22)	129 745 (19.6)
Falls or fragility fractures, *n* (%)	2056 (2.6)	32 199 (4.9)
New home care package, *n* (%)	748 (0.9)	N/A
Care home admission, *n* (%)	343 (0.4)	5381 (0.8)
All-cause mortality, *n* (%)	2560 (3.3)	25 216 (3.8)
Composite outcome*, *n* (%)	4800 (6.1)	56 408 (8.5)

Key: IMD, index of multiple deprivation- The English IMD was applied to the Development cohort and the Welsh IMD was applied to the external validation cohort. a) Care home admission only counted where there was no pre-existing admission before the study period; b) Home care package only counted where there was no pre-existing home care package before the study period; *Composite outcome does not include new home care package in external validation data. N/A, data not available.

evidence of miscalibration (calibration slope = 1.104, 95% CI: 1.093,1.113) with the calibration slope indicating the spread of predicted risks may be too narrow. The underestimation of observed risk was apparent over the whole spectrum of predicted risk (Figure 1b).

The median eFI2 score was 0.066 (95% CI: 0.036, 0.119) in the external validation cohort. The reference standard cut-point of 0.25, between robust and mild frailty, was mapped to an eFI2 score of 0.0857 using the CARE75+ data. Cut-points to distinguish between mild, moderate and severe frailty were 0.1624 and 0.2392. Using these cut-points, 402 427 (60.9%) were identified as robust, 169 301 (25.6%) as mild frailty, 63 019 (9.5%) as moderate frailty and 25 670 (3.9%) as severe frailty. The risk of the composite outcome increased over the frailty categories (log-rank test $P < 0.001$) (Supplementary Materials, Appendix 2). Similarly, the individual risks of falls, care home admission and mortality increased with increasing frailty (Supplementary Materials, Appendix 4).

The original eFI had moderate discrimination in the external validation cohort (C-index = 0.687, 95% CI: 0.684, 0.689, Nagelkerke's $R^2 = 0.050$).

Implementation

An example of how to calculate the eFI2 score, the eFI2 frailty category and the predicted risk of the composite outcome at 12 months is shown in Supplementary Materials, Appendix 3.

Discussion

We have developed and externally validated the eFI2, which demonstrates good predictive performance for the composite of four key frailty outcomes. We have defined new frailty categories using a reference standard FI measure. While there was evidence that the eFI2 underestimated the risk of frailty in the external validation cohort, the eFI2 had better discrimination for the composite outcome than the original eFI.

The study is strengthened by the use of large population-based cohorts, with external validation to examine potential overfitting of the model and assess generalisability. We ensured high ascertainment of predictors and outcomes by developing, testing and implementing an inclusive list of SNOMED CT codes. The mapping from SNOMED CT to Read v2 suggests that the eFI2 could be successfully translated to different coding systems. However, some variation in coding systems is inevitable and likely contributed to variation in the prevalence for some predictors and outcomes.

The eFI2 underestimated the observed risk in the external validation cohort. Several factors may have contributed to this. Firstly, the external validation cohort included care home residents and people in receipt of a new home care package. These patients were excluded from the development cohort, reducing incidence of the composite outcome from 7.4% to 6.1%. Conversely, although we included people in receipt of home care packages in the external validation cohort, it was not possible to ascertain the home care package outcome in SAIL, which may have reduced the baseline hazard of the composite outcome. Secondly, the external validation cohort had a greater incidence of injurious falls, even after accounting for the inclusion of patients with previous care home admissions or people in receipt of a home care package. The higher incidence might represent a real difference in the populations, differences in coding, or use of different coding ontologies. Lastly, the prevalence of a small number of predictors varied considerably between cohorts.

Table 2. Prevalence of deficit variables included as predictors in the eFI2.

Deficit	Prevalence n(% of cohort)	
	Development cohort *N* = 81 685	External validation cohort *N* = 660 417
Alcohol		
Harmful intake	2943 (3.7)	4714 (0.7)
Higher intake	883 (1.1)	686 (0.1)
Lower intake	5613 (7.1)	11 231 (1.7)
Previous harmful/higher	25 (0.03)	90 (0.01)
Zero	5303 (6.7)	1247 (0.2)
Missing	63 993 (81.3)	642 449 (97.3)
Activity limitation	1293 (1.8)	5646 (0.9)
Atrial fibrillation	11 048 (14.0)	59 098 (8.9)
Cancer	20 642 (26.2)	134 167 (20.3)
BMI		
Underweight	1424 (1.8)	14 856 (2.2)
Recommended	23 695 (30.1)	180 164 (27.3)
Overweight	28 883 (36.7)	224 119 (33.9)
Obese	21 151 (26.9)	171 436 (25.96)
Missing	3607 (4.6)	69 392 (10.5)
Cognitive impairment	25 890 (32.9)	6644 (1.0)
COPD	10 230 (13.0)	77 849 (11.8)
Dementia	7003 (8.9)	18 870 (2.9)
Dressing or grooming problems	10 662 (13.5)	<10
Environment problems	2577 (3.3)	12 249 (1.9)
Falls	16 163 (20.5)	106 839 (16.2)
Fracture	18 255 (23.2)	157 004 (23.8)
Fragility fracture	13 947 (17.7)	79 033 (12.0)
Heart failure	8386 (10.6)	76 939 (11.7)
Housebound	7076 (9.0)	74 800 (11.3)
Hypotension or syncope	13 409 (17.0)	48 961 (7.4)
Liver problems	1071 (1.4)	3400 (0.5)
Medication management problems	432 (0.5)	<10
Memory concerns	844 (1.1)	25 552 (3.9)
Mobility problems	13 796 (17.5)	8894 (1.3)
Motor Neuron Disease	81 (0.1)	251 (0.04)
Palliative care	4435 (5.6)	5451 (0.8)
Parkinsonism or tremor	2285 (2.9)	20 670 (3.1)
Peptic ulcer disease	922 (1.2)	4941 (0.7)
Peripheral vascular disease	10 020 (12.7)	44 496 (6.7)
Polypharmacy[a]		
0–4 medications	45 015 (57.2)	392 508 (59.4)
5–9 medications	23 132 (29.4)	202 158 (30.6)
10+ medication	10 613 (13.5)	65 751 (10.0)
Requirement for care	4794 (6.1)	23 427 (3.5)
Respiratory disease	8285 (10.5)	58 989 (8.9)
Seizures	1922 (2.4)	15 238 (2.3)
Self-harm	566 (0.7)	1601 (0.2)
Skin ulcer	8919 (11.3)	75 506 (11.4)
Social vulnerability	8956 (11.4)	186 887 (28.3)
Stroke	9776 (12.4)	48 091 (7.3)
Smoking		
None, Ex or missing	57 225 (72.7)	473 530 (71.7)
Current	21 535 (27.3)	186 887 (28.3)
Transient ischaemic attack	4577 (5.8)	29 018 (4.4)
Weight loss	4998 (6.3)	49 466 (7.5)

COPD = Chronic obstructive pulmonary disease, BMI = Body mass index. More information on the definitions of the predictors in available in Supplementary Materials, Appendix 1. [a]Number of medications from different BNF sub-subchapters (level 3) prescribed in previous 90 days.

Table 3. eFI2 model coefficients and transformed coefficients (n = 36)

Predictor	Coefficient	Transformed coefficient
Activity limitation	0.15284	0.018
Alcohol, harmful intake	0.23107	0.027
Alcohol, missing	0.13175	0.016
Alcohol, previous harmful/higher	1.36434	0.162
Atrial fibrillation	0.13025	0.015
Cancer	0.2406	0.029
Cognitive impairment	0.10985	0.013
COPD	0.11683	0.014
Dementia	0.41715	0.049
Dressing or grooming problems	0.05422	0.006
Environment problems	0.11886	0.014
Falls (history of)	0.62743	0.074
Fracture	0.07353	0.009
Fragility fracture	0.17425	0.021
Heart failure	0.11086	0.013
Housebound	0.33254	0.039
Hypotension or syncope	0.18253	0.022
Liver problems	0.23787	0.028
Medication management problems	0.32125	0.038
Memory concerns	0.11915	0.014
Mobility problems	0.46836	0.056
Motor neuron disease	0.35347	0.042
BMI, missing	0.25318	0.030
BMI, underweight	0.4417	0.052
Palliative care	0.5145	0.061
Parkinsonism or tremor	0.03537	0.004
Peptic ulcer disease	0.05427	0.006
Peripheral vascular disease	0.02672	0.003
Polypharmacy, 5–9 medications	0.32301	0.038
Polypharmacy, 10+ medications	0.50801	0.060
Requirement for care	0.21428	0.025
Respiratory disease	0.01049	0.001
Seizures	0.02885	0.003
Self-harm	0.00900	0.001
Skin ulcer	0.21935	0.026
Smoker current	0.10291	0.012
Social vulnerability	0.23585	0.028
Stroke	0.10565	0.013
Transient ischaemic attack	0.02305	0.003
Weight loss	0.19256	0.023

BMI = body mass index, COPD = chronic obstructive pulmonary disease. More information on the definitions of the predictors in available in Supplementary Box 1. A worked example for calculating a hypothetical patient's eFI2 score, frailty category and predicted risk is shown in Appendix 3.

Table 4. Model performance statistics

Statistic	Internal validation[d] Statistic	External validation Statistic (95% CI)
Calibration slope[a]	1.000	1.104 (1.093, 1.113)
Concordance (C-statistic)[b]	0.803	0.723 (0.721, 0.725)
Nagelkerke's R^2	0.0971	0.064
E/O[c]	0.987	0.443

[a]The calibration slope represents the spread of expected versus observed incidence, with a slope > 1 indicating that the spread of predicted risk are too narrow and a slope < 1 that the spread is too extreme [27]. [b]The C-index ranges from 0 to 1 with a value of 0.5 suggesting the model predicts the composite outcome no better than chance, and 1 being indicative of perfect discrimination such that everyone in the data with the composite outcome have higher predicted risk than all those without the outcome [27]. [c]The E/O ratio compares the one year incidence (expected) to the incidence in reality (observed). An E/O ratio of 1 indicates that the predicted one-year incidence of the composite outcome is equal to the observed incidence of the composite outcome. Where the ratio is under 1, this suggests that the model under-estimates the incidence compared to what happened in reality (and vice vera) [27]. [d]Bootstrapped statistics (50 bootstraps).

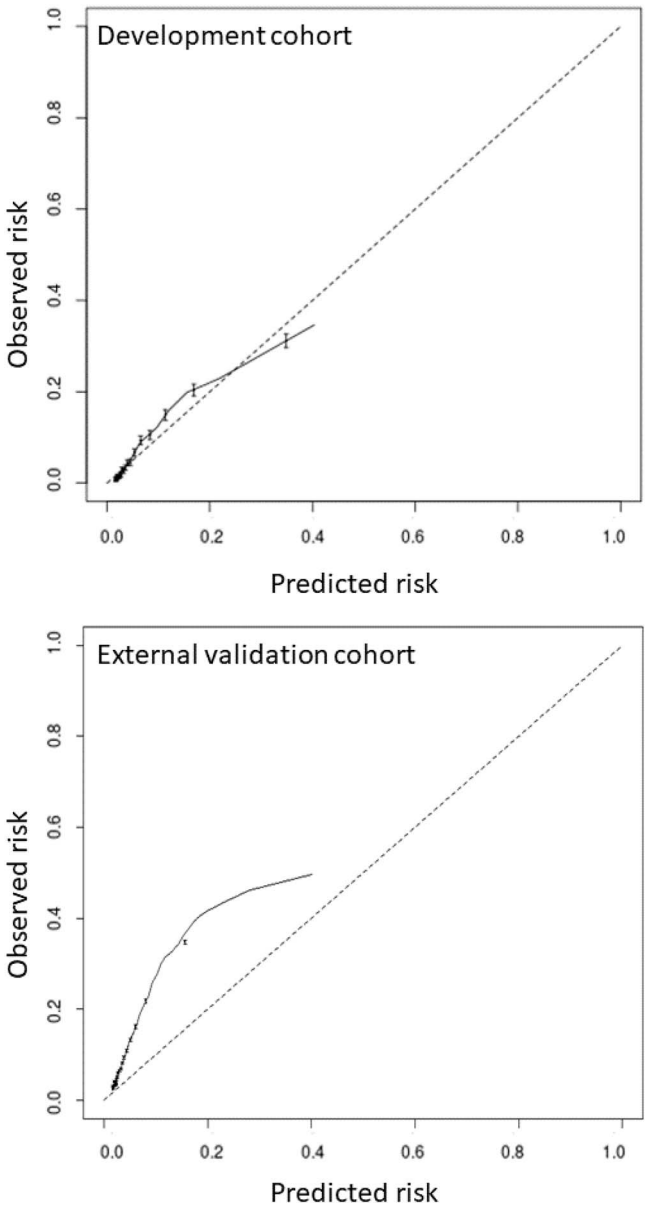

Figure 1. Calibration curves for performance of the eFI2 model at one year. Groups represent 20ths of the linear predictor.

The prevalence of cognitive impairment in particular was significantly higher in the development cohort where primary care incentivisation schemes increased recording.

We combined four frailty-related outcomes and used this as a single composite outcome. As a marker of independence, we were able to include local authority organised/funded care as an outcome in the development cohort. However, it was not possible to ascertain informal care or privately organised/funded care, which means we have not perfectly captured loss of independence. While the size of the association with the predictors is likely to vary for each outcome, they are the key outcomes that best represent the overall state of frailty.

Our novel approach to eFI2 development, with use of weighted predictor variables, time-constraints and updated frailty categories has resulted in good prognostic performance for our composite outcome. Although the main use of eFI2 should be for population-level risk stratification to identify groups of individuals who should be considered for frailty-related interventions, in primary care batch coding should be avoided and clinical judgement should be used to confirm a formal diagnosis of mild, moderate or severe frailty before adding this to the clinical record, aligned with previous NHS guidance on eFI implementation [32]. We mapped from a research standard frailty index to define frailty categories, although alternative approaches could have been used, for example to employ decision curve analysis or to work with clinical and patient reference groups to derive risk thresholds [33].

The robust prognostic performance of the eFI2 and the improved classification of frailty categories could support development of improved care pathways for older people living with frailty. Our use of global standard SNOMED CT nomenclature will support external validation of the eFI2 internationally as a necessary step prior to international implementation, with potential for global scientific, practice and health policy impact, given the contemporary international interest in this area [34].

Potential next steps include decision curve analysis, or decision analytic modelling to assess how model predictions might lead to improved patient and system-level outcomes, including the economic perspective [33]. Comparison of the eFI2 frailty categories to those derived from clinical measures of frailty would add further evidence on overall validity, along with further validation to assess performance across different demographics.

Conclusion

The eFI2 demonstrates robust prediction for key frailty-related outcomes, with improved discrimination for a composite outcome of falls, care home admission and mortality compared with the original eFI. We anticipate that the eFI2 will support development and implementation of new models of primary care for older people living with frailty in the United Kingdom, with future impact extending into other key areas including perioperative care, oncology and cardiovascular care. Our use of novel methodology to develop and validate the eFI2 will advance the field of frailty-related research internationally, setting a new methodological standard for international research. Similar clinical impact could be anticipated internationally, following future external validation of eFI2 in international settings prior to implementation.

Acknowledgements: This study is based [in part] on data from Connected Bradford (REC 18/YH/0200 & 22/EM/0127). The data is provided by the citizens of Bradford and district, and collected by the NHS, Department for Education (DfE) and other organisations as part of their care and support. The interpretation and conclusions contained

in this study are those of the authors alone. The NHS, DfE and other organisations do not accept responsibility for inferences and conclusions derived from their data by third parties. The interpretation and conclusions contained in this study are those of the authors alone. This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who make anonymised data available for research. We are extremely grateful to the members of our Frailty Oversight Group, NIHR ARCYH Implementation Advisory Group, and Frailty Clinical Network, NHS England—London Region for their expert guidance across the project.

Supplementary Data: Supplementary data is available at *Age and Ageing* online.

Declaration of Conflicts of Interest: None.

Declaration of Sources of Funding: This work was funded by the National Institute for Health Research Applied Research Collaboration Yorkshire & Humber (NIHR ARC YH) and supported by funding from the NIHR ARC North Thames. K.B. is supported by NIHR ARC YH and the NIHR Leeds Biomedical Research Centre. D.N. is funded by a Fellowship award from the Alzheimer's Society, UK (grant number 584). C.A. was funded by an NIHR Development and Skills Enhancement (DSE) Award. A.C. is funded by an NIHR Research Professorship and supported by NIHR ARC YH, the NIHR Leeds Biomedical Research Centre and Health Data Research UK, an initiative funded by UK Research and Innovation Councils, NIHR and the UK devolved administrations and leading medical research charities. The views expressed are those of the authors, and not necessarily those of the NIHR or Department of Health and Social Care.

Ethics: Connected Bradford has ethical approval for a research database (REC 18/YH/0200 & 22/EM/0127) and this study did not require independent ethical approval. All SAIL Databank projects are reviewed for approval by the Information Governance Review Panel (IGRP), separate ethical approval is not required for the use of anonymised SAIL data.

Research data transparency and availability statement: The eFI2 model equation and associated code lists used to define variables are available from the corresponding author for research use. We aim to make eFI2 available to suppliers of UK electronic health record systems, risk stratification software, and for use in NHS policy and commissioning under the terms of an agreed licence agreement. Organisations wishing to use eFI2 for commercial purposes should contact the corresponding author directly.

References

1. Prof Chris Whitty. *Chief Medical Officer's Annual Report 2023: Health in an Ageing Society*. 2023. https://www.gov.uk/government/publications/chief-medical-officersannual-report-2023-health-in-an-ageing-society
2. Clegg A, Young J, Iliffe S *et al.* Frailty in elderly people. *Lancet* 2013;**381**:752–62. https://doi.org/10.1016/S0140-6736(12)62167-9.
3. Nikolova S, Heaven A, Hulme C. Social care costs for community-dwelling older people living with frailty. *Health Social Care Commun* 2022;**30**:804–11. https://doi.org/10.1111/hsc.13450.
4. Crocker TF, Ensor J, Lam N *et al.* Community based complex interventions to sustain independence in older people: Systematic review and network meta-analysis. *BMJ* 2024;**384**:e077764. https://doi.org/10.1136/bmj-2023-077764.
5. Ho L, Malden S, McGill K. Complex interventions for improving independent living and quality of life amongst community-dwelling older adults: A systematic review and meta-analysis. *Age Ageing* 2023;**52**:afad132. https://doi.org/10.1093/ageing/afad132.
6. Kim DH, Rockwood K. Frailty in older adults. *N Engl J Med* 2024;**391**:538–48. https://doi.org/10.1056/NEJMra2301292.
7. Negm AM, Kennedy CC, Thabane L. Management of frailty: A systematic review and network meta-analysis of randomized controlled trials. *J Am Med Dir Assoc* 2019;**20**:1190–8. https://doi.org/10.1016/j.jamda.2019.08.009.
8. Bergmann MM, Calle EE, Mervis CA *et al.* Validity of self-reported cancers in a prospective cohort study in comparison with data from state cancer registries. *Am J Epidemiol* 1998;**147**:556–62. https://doi.org/10.1093/oxfordjournals.aje.a009487.
9. Quint JK, Müllerova H, DiSantostefano RL. Validation of chronic obstructive pulmonary disease recording in the clinical practice research datalink (CPRD-GOLD). *BMJ Open* 2014;**4**:e005540. https://doi.org/10.1136/bmjopen-2014-005540.
10. Morley JE, Vellas B, van Kan. Frailty consensus: A call to action. *J Am Med Dir Assoc* 2013;**14**:392–7. https://doi.org/10.1016/j.jamda.2013.03.022.
11. Centre for Perioperative Care and British Geriatrics Society. *Guideline for Perioperative Care for People Living with Frailty Undergoing Elective and Emergency Surgery*. 2021 [cited 2023 29/9/2023]. https://www.cpoc.org.uk/sites/cpoc/files/documents/2021-09/CPOC-BGS-Frailty-Guideline-2021.pdf.
12. Richter D, Guasti L, Walker D *et al.* Frailty in cardiology: Definition, assessment and clinical implications for general cardiology. A consensus document of the Council for Cardiology Practice (CCP), Association for Acute Cardio Vascular Care (ACVC), Association of Cardiovascular Nursing and Allied Professions (ACNAP), European Association of Preventive Cardiology (EAPC), European heart rhythm association (EHRA), council on Valvular heart diseases (VHD), council on hypertension (CHT), Council of Cardio-Oncology (CCO), working group (WG) aorta and peripheral vascular diseases, WG e-cardiology, WG thrombosis, of the European Society of Cardiology, European primary care cardiology society (EPCCS). *Eur J Prev Cardiol* 2021;**29**:216–27. https://doi.org/10.1093/eurjpc/zwaa167.
13. Ethun CG, Bilen MA, Jani AB *et al.* Frailty and cancer: Implications for oncology surgery, medical oncology, and radiation oncology. *CA Cancer J Clin* 2017;**67**:362–77. https://doi.org/10.3322/caac.21406.

14. Clegg A, Bates C, Young J. Development and validation of an electronic frailty index using routine primary care electronic health record data. *Age Ageing* 2016;**45**:353–60. https://doi.org/10.1093/ageing/afw039.
15. NHS England. *NHS England Standard General Medical Services Contract 2017/18*, 2017/2018.
16. NHS England. *Network Contract Directed Enhanced Service: Contract Specification 2022/23-PCN Requirements and Entitlements*, 2022.
17. Rockwood K, Mitnitski AJTJOGSABS, Sciences M. Frailty in relation to the accumulation of deficits. 2007;**62**:722–7.
18. Sohal K, Mason D, Birkinshaw J. Connected Bradford: A whole system data linkage accelerator. *Wellcome Open Res* 2022;**7**:26. https://doi.org/10.12688/wellcomeopenres.17526.1.
19. SAIL Databank. *SAIL Databank- about us*. 2022. https://saildatabank.com/.
20. van der Windt DA, Jordan J, Cottrell E *et al. Predictors of Unplanned Hospital Admissions, Mortality and Poor Quality of Life in Older People Who Are Identified as Frail in Primary Care. PROSPERO CRD42019111775*. 2019. http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42019111775.
21. NHS England. *2018/19 General Medical Services (GMS) Contract Quality and Outcomes Framework (QOF)*, 2018.
22. Riley RD, Ensor J, Kym ES I. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;**368**:m441. https://doi.org/10.1136/bmj.m441.
23. Rubenstein LZ. Falls in older people: Epidemiology, risk factors and strategies for prevention. *Age Ageing* 2006;**35**: ii37–41. https://doi.org/10.1093/ageing/afl084.
24. Riley RD, Collins GS, Ensor J. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. *Stat Med* 2022;**41**:1280–95. https://doi.org/10.1002/sim.9275.
25. J Goeman, R Meijer, and N. Chaturvedi. *L1 and L2 Penalized Regression Models*. 2022 cited 2023; Available from: L1 and L2 Penalized Regression Models.
26. Searle SD, Mitnitski A, Gahbauer EA *et al.* A standard procedure for creating a frailty index. *BMC Geriatr* 2008;**8**:24. https://doi.org/10.1186/1471-2318-8-24.
27. Collins GS, Dhiman P, Ma J. Evaluation of clinical prediction models (part 1): From development to external validation. *BMJ* 2024;**384**:e074819. https://doi.org/10.1136/bmj-2023-074819.
28. Royston P, Altman DG. External validation of a cox prognostic model: Principles and methods. *BMC Med Res Methodol* 2013;**13**:1–15. https://doi.org/10.1186/1471-2288-13-33.
29. Heaven A, Brown L, Young J *et al.* The community ageing research 75+ study (CARE75+): An experimental ageing and frailty research cohort. *BMJ Open* 2019.
30. Marshall A, Nazroo J, Tampubolon G *et al.* Cohort differences in the levels and trajectories of frailty among older people in England. *J Epidemiol Community Health* 2015; **69**:316–21. https://doi.org/10.1136/jech-2014-204655.
31. Heaven A, Brown L, Foster M *et al.* Keeping it credible in cohort multiple randomised controlled trials: The community ageing research 75+(CARE 75+) study model of patient and public involvement and engagement. *Res Involve Engage* 2016;**2**:1–12. https://doi.org/10.1136/jech-2014-204655.
32. England, N. Identifying frailty. 2017/18. https://www.england.nhs.uk/ourwork/clinical-policy/older-people/frailty/frailty-riskidentification/.
33. Vickers AJ, van Calster, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnost Prognost Res* 2019;**3**:18. https://doi.org/10.1186/s41512-019-0064-7.
34. Orkaby AR, Callahan KE, Driver JA *et al.* New horizons in frailty identification via electronic frailty indices: Early implementation lessons from experiences in England and the United States. *Age Ageing* 2024;**53**. https://doi.org/10.1093/ageing/afae025.

Received 28 June 2024; editorial decision 17 February 2025