Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Incremental predictive value of cystatin C for next-visit care needs in older adults: a longitudinal CHARLS study

Yanche Lu
Wenzhou Medical University
Ruijie Lin
Wenzhou Medical University
Qian Yu
Wenzhou Medical University
Qiuyu Zheng
Wenzhou Medical University
Lianmin Ye
ylmwork88@126.com
Wenzhou Medical University

Research Article

Keywords: Cystatin C, Care need, Activities of daily living, Prediction model, Incremental value, CHARLS

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11376472/v1

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

Additional Declarations: No competing interests reported.

Incremental predictive value of cystatin C for next-visit care needs in older adults: a longitudinal CHARLS study

Yanche Lu[1,2†], Ruijie Lin[1,2†], Qian Yu[1,2], Qiuyu Zheng[1,2] and Lianmin Ye[1,2*]

[1] Wenzhou Medical University, Wenzhou, Zhejiang 325035, China

[2] Department of Clinical Medicine, Wenzhou Medical University Renji College, Wenzhou, Zhejiang 325035, China

† These authors contributed equally to this work.

* Corresponding author: Lianmin Ye. Email: ylmwork88@126.com.

Abstract

Background: We assessed the incremental predictive value of cystatin C (CysC) for next-visit care needs in older adults.

Methods: Multinomial ridge models were developed in 5,854 person-windows from 4,385 China Health and Retirement Longitudinal Study (CHARLS) participants aged 60–100 years without baseline basic activities of daily living (BADL) assistance needs. Evaluation included 1,901 participants from separate communities in 2018–2020, using historical 2015 laboratory and physical measurements. Outcomes were survival without assistance, survival with assistance, and death. A reference model containing clinical, functional, and laboratory predictors (B1) was compared with B1 plus CysC (B2). Exploratory extensions added grip strength, C-reactive protein, and hemoglobin, applied development-only calibration, and compared 10%, 20%, and 30% assessment capacities using 1,000 paired community bootstrap samples. The holdout had been used previously.

Results: Evaluation included 98 care-need states and 69 deaths. For the original, separately tuned B1/B2 models, areas under the receiver operating characteristic curve (AUCs) were 0.7205/0.7214 (difference, 0.0009; 95% confidence interval [CI], −0.0096 to 0.0126); the temperature-calibrated binary Brier difference was −0.001345 (−0.002304 to −0.000288). Beyond the enhanced reference, raw AUC increments were 0.0007 (−0.0066 to 0.0072) at an inverse regularization strength of 0.03 and −0.0012 (−0.0088 to 0.0052) at 0.003.

Simple intercept calibration retained a small reduction in Brier score at 0.03; its interval crossed zero at 0.003. Neither condition showed clear net-benefit or fixed-capacity care-capture gains.

Conclusions: Historical CysC yielded small, regularization-dependent reductions in probability error beyond clinical and muscle-strength measures. Gains in selecting participants for comprehensive assessment remained uncertain and require independent validation.

Keywords: Cystatin C; Care need; Activities of daily living; Prediction model; Incremental value; CHARLS

1. Background

A new need for help with basic activities of daily living (BADL), such as dressing, bathing, or toileting, changes an older adult's daily care requirements. Healthy aging and long-term care are priorities in China's response to population aging [1,2]. Projections based on the China Health and Retirement Longitudinal Study (CHARLS) indicate that population aging can increase the number of people needing care despite declining age-specific care needs [3]. Anticipating this transition among currently independent older adults may help community health services prioritize comprehensive geriatric assessment and follow-up, including evaluation of nutritional needs. Geriatric nutrition guidance recommends routine malnutrition screening and individualized nutritional care as part of a comprehensive, multidisciplinary approach [4]. Prediction models in Chinese populations already combine demographic characteristics, health conditions, and functional information to estimate disability or dependency [5–7]. The clinical question is whether an existing cystatin C (CysC) result adds predictive information beyond these measures when identifying people who may need comprehensive geriatric assessment.

CysC is a blood-based measure of kidney function with potential relevance to functional aging. CysC-based equations provide alternatives for estimating glomerular filtration rate (GFR), and equations combining creatinine with CysC improve estimation of measured GFR [8,9], although CysC also has nonfiltration determinants [10]. In CHARLS and the US Health and Retirement Study, CysC-related kidney function measures were associated with frailty progression and decline in physical function [11]. Other CHARLS analyses linked creatinine- and CysC-based indices to subsequent frailty and activities of daily living (ADL) disability [12] and to heterogeneous ADL disability trajectories [13]. One disability prediction model included CysC among its candidate laboratory predictors [7]. CysC–creatinine scores have also been evaluated for prevalent frailty across populations, and a creatinine–CysC-derived index has been examined for sarcopenia

incidence or progression [14,15]. It remains unclear whether historical CysC adds predictive information for next-visit BADL assistance beyond directly measured strength and existing clinical information.

A biomarker may change how well a model ranks individuals or estimates their outcome probabilities. At a decision threshold, it may also alter the balance between useful and unnecessary assessments. Discrimination, calibration, prediction error, and decision net benefit capture these distinct aspects of prediction [16–19]. We evaluated the incremental predictive value of CysC beyond a reference model containing demographic, functional, disease-related, creatinine, metabolic, and systolic blood pressure information. Models were developed in two CHARLS visit windows and evaluated in a later window from separate communities. Death was retained as a distinct outcome state, and the target was being alive and needing BADL assistance at the next visit. Comparisons at common regularization parameters assessed how the estimated increment depended on shrinkage. This design evaluated the additional information available for next-visit care assessment within CHARLS over the observed visit intervals. Exploratory analyses extended the reference model with grip strength, C-reactive protein (CRP), and hemoglobin. They also compared simple calibration methods applied symmetrically to paired models and care capture at equal assessment capacities.

2. Methods

2.1. Data source and study design

CHARLS is a longitudinal household survey of Chinese residents aged 45 years and older and their spouses [20]. We used questionnaire, physical examination, blood, and survival-status data from the 2011, 2013, 2015, 2018, and 2020 waves. Development combined the 2013–2015 and 2015–2018 windows in development communities; primary validation used the 2018–2020 window in holdout communities (Figure 1). Participants could contribute up to two equally weighted development person-windows and one primary validation record. Earlier windows in holdout communities were examined in secondary analyses.

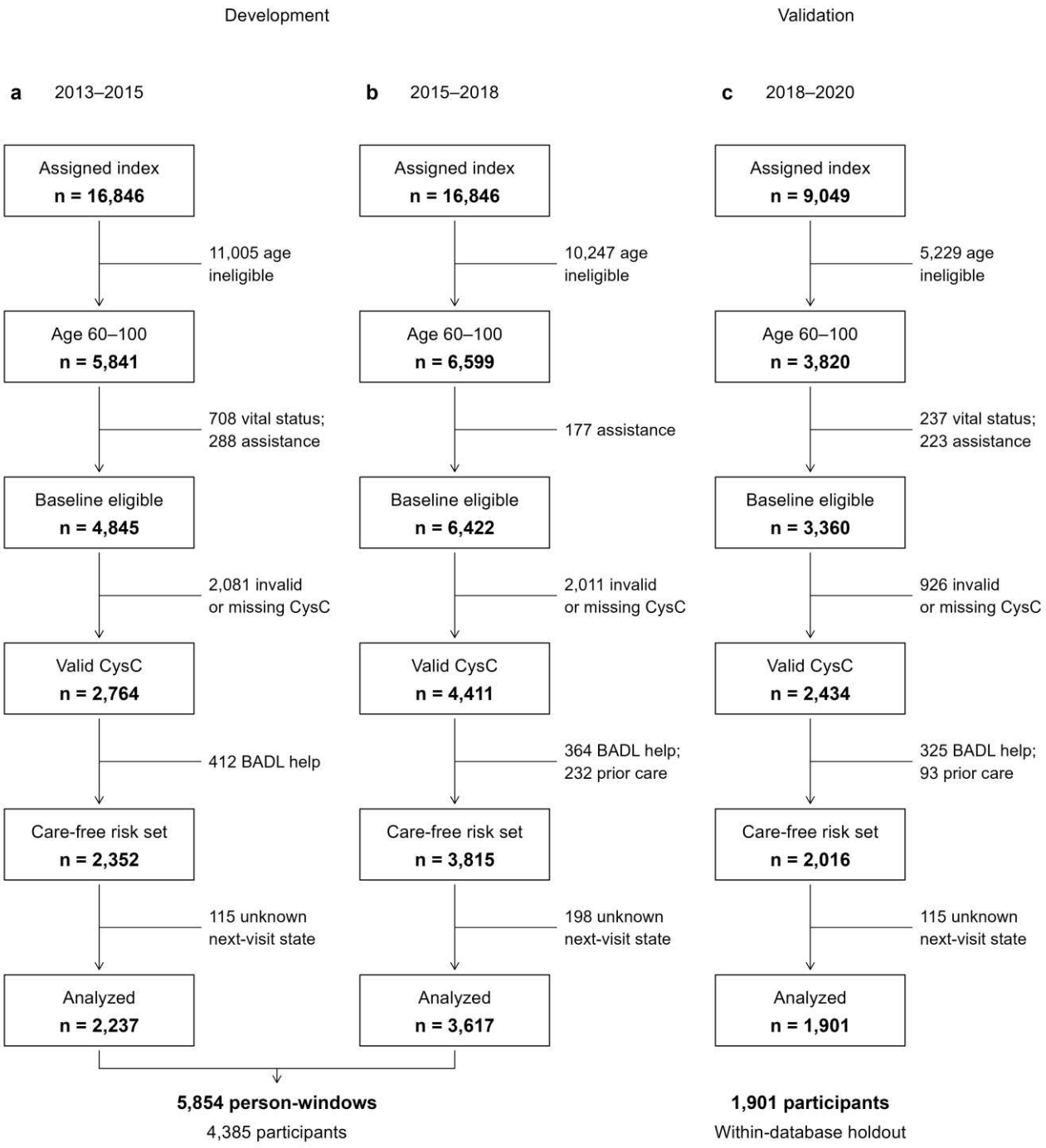

Figure 1. Participant selection for model development and validation. a,b, Sequential selection for the 2013–2015 and 2015–2018 development windows. c, Selection for the 2018–2020 within-database validation window. Boxes show retained counts; unboxed side labels show mutually exclusive exclusions between successive stages. Assigned-index counts are community-allocation counts from a cross-wave index of 26,052 people, rather than wave-specific interview totals; 157 people could not be assigned a fixed community. Vital-status exclusions lacked confirmed survival at both designated blood-measurement and prediction waves. Participants were excluded if their response-assistance category was outside the permitted categories 1–3 at either wave. The care-free risk set required confirmed absence of basic activities of daily living (BADL) assistance need at both waves and no previously recorded care-need status. Cystatin C (CysC) came from 2011/12 blood measurements for the first window and 2015-wave measurements for the other two. Analyzed participants had a classifiable next-visit state: alive without BADL help, alive needing help, or death; death was retained as a distinct state. The numbers of unknown outcomes were 115, 198, and 115. Development comprised 5,854 person-windows from 4,385 participants, including 1,469 repeat contributors. Validation comprised 1,901 participants: 1,734 alive without BADL assistance needs, 98 alive needing BADL assistance, and 69 who had died. The holdout communities were also used for historical-window evaluation. Full sequential exclusions are provided in Table S7.

Clinical and functional predictors were obtained at the prediction-start interview. The 2013 start used 2011/2012 blood measurements and 2013 blood pressure. The 2015 start used blood and physical measurements from the 2015 wave. The 2018 start combined 2018 interviews with blood measurements, blood pressure, body mass index, and exploratory grip-strength measurements from the 2015 wave. Historical measurements were not replaced with values from other waves. The released blood files did not contain participant-specific collection dates, and the 2015-wave interview records provided the year and month but not the day. The within-wave order of blood collection and interviews therefore could not be established for the 3,617 development person-windows in the 2015–2018 window. Sixteen of these person-windows had interviews in 2016. The 2011–2013 window was excluded because the order of baseline interviews and blood collection could not be reliably established. Nominal visit interval, two or three years, was included as a predictor. The estimand was status at the next visit, not risk over a uniform two-year horizon.

In 2011, CysC was measured in all blood providers aged 70 years and older and in a random subsample of those younger than 70. Blood collection and transport conditions also differed between 2011 and 2015 [21,22]. The target population was therefore restricted to eligible participants whose CysC had been measured in the required wave. Analyses were unweighted, and the estimated risks apply to this selected sample.

2.2. Participants and sample allocation

Eligible participants were aged 60–100 years at the prediction start, had a valid CysC measurement in the required blood wave, and had consistent evidence of survival and confirmed absence of BADL assistance needs at both the blood and prediction-start waves. Interview-assistance frequency at both assessments had to be never, some of the time, or most or all of the time; fully proxy and unknown responses were excluded. Any earlier recorded state requiring assistance also led to exclusion. Nonparticipation or unknown status in an earlier wave was not classified as negative. Inclusion required ascertainable status at the required baseline assessments and no recorded prior positive state. The main models included participants with a known next-visit three-state outcome; unknown outcomes were retained in the follow-up record. B0, B1, and B2 used identical analysis samples, and a valid CysC measurement was also required for B0 and B1. Study size was determined by all eligible records in the available waves and fixed community partitions.

Communities were allocated within province-by-urban/rural strata using deterministic SHA-256 ordering of a fixed salt and community identifier. Approximately 70% of communities in each stratum entered development

and the remainder entered holdout; singleton strata entered development (Additional file 1: Supplementary Methods S2). The allocation preceded the analysis eligibility filters. Stable participant and household identifiers kept repeated records and household members within the same partition, and all training folds were grouped by community. Development and primary validation had no participant, household, or community overlap. The within-database holdout had been examined in earlier analyses before the current primary model protocol was frozen on 7 October 2026. The analysis chronology is provided in Supplementary Methods S9. Reporting was informed by TRIPOD+AI and TRIPOD-Cluster [23,24]. An item-by-item TRIPOD+AI reporting checklist is provided in Supplementary Table S8.

The CHARLS blood biomarker collection protocol was approved by the Peking University institutional review board (IRB00001052-11014), and participants provided written informed consent [22]. The secondary-analysis ethics statement is provided in the Ethics approval and consent to participate section.

2.3. Outcome definition

The outcome comprised three mutually exclusive states: alive without BADL assistance needs (state 0), alive and needing help with at least one BADL activity or unable to perform it (state 1), and confirmed death by the next visit (state 2). BADL activities were dressing, bathing, eating, getting into or out of bed, toileting, and continence. Difficulty while remaining independent did not meet the care-need definition [25]. A valid assistance response for any one activity established state 1; state 0 required ascertainable negative responses for all six activities. Before 2020, missing items were resolved only using valid questionnaire skip rules, without overriding conflicting difficulty responses. Those skip-based rules were not applied to 2020 data.

Conflicting death and survival records, implausible death timing, missing interviews, or insufficient BADL information were classified as unknown. Death remained a separate state in the model. The primary AUC treated state 1 as positive and states 0 and 2 as negative for state 1; a conditional AUC among survivors was also reported. This definition does not capture transient assistance needs between visits or unrecorded care needs before death, nor does it estimate the cumulative incidence of a first care-need event.

2.4. Predictors and measurements

Three nested information blocks were evaluated in the same participants. B0 included age, sex, two education indicators, hypertension, diabetes, number of BADL difficulties, difficulty walking 100 m, the 10-item Center

for Epidemiologic Studies Depression Scale (CES-D 10) score, fair or worse self-rated health, number of instrumental activities of daily living (IADL) difficulties, number of mobility difficulties, creatinine, most recently available body mass index, smoking at the blood wave, non-high-density lipoprotein (non-HDL) cholesterol, glycated hemoglobin, and nominal visit interval. B1 added the most recently available systolic blood pressure; B2 added CysC. The blocks contained 18, 19, and 20 original variables, respectively. Predictors were selected for clinical relevance and availability, without univariable significance screening. B0 included laboratory information.

CysC was recorded in mg/L, with valid values of 0.2–5, and creatinine in mg/dL, with valid values of 0.1–20; both were transformed using the natural logarithm. Non-HDL cholesterol was calculated as total minus HDL cholesterol, with negative values set to missing; nonpositive glycated hemoglobin values were also set to missing. Systolic blood pressure required three valid integer readings of 70–260 mmHg, paired diastolic readings of 40–150 mmHg, and systolic pressure exceeding diastolic pressure in each pair. The mean of the second and third systolic readings was used; failure of any rule made systolic pressure missing for that wave. Body mass index was calculated from valid height and weight measurements and retained within 12–60 kg/m². Original fields, derivation rules, and measurement waves are provided in Additional file 1.

Exploratory extensions retained B1 and added continuous grip strength (G), then natural-log CRP and hemoglobin (S), and finally CysC (SC). Grip strength, in kg, was the maximum valid value across two readings in each hand; codes 993/999 and values outside 0–100 kg were missing, while a recorded zero was retained. The upper limit was a study quality-control rule. Positive hemoglobin (g/dL) and CRP (mg/L) values from the released data were retained without outcome-guided truncation; hemoglobin entered the model linearly. Grip measurements came from 2013/2015/2015 and blood measurements from 2011/2015/2015 for the three windows. All comparisons used the same participants, imputation learned from training data, and fixed missingness indicators. Body mass index was already included in B1. Field definitions, noncompletion reasons, and extreme values are provided in Supplementary Methods S10 and Table S13.

2.5. Model development and calibration

Community-grouped nested cross-validation, with five outer and four inner folds, compared multinomial ridge logistic regression, spline ridge logistic regression, and histogram gradient boosting. The ridge grid used C=0.003, 0.03, and 0.3; spline ridge used C=0.03 and 0.3; and boosting compared models with three and seven

leaves. C denotes inverse regularization strength, so smaller values imply stronger shrinkage. Algorithm-family selection used the care-need area under the receiver operating characteristic curve (AUC) from B2 outer-fold predictions. For AUC differences of 0.005 or less, multiclass log loss, model complexity, and penalty strength were considered in that order. All information blocks used the selected family, with separate tuning using identical community folds. Final C values were selected separately for each information block by five-fold community-grouped cross-validation on the full development sample, using pooled out-of-fold predictions and the same selection rule. Algorithm-family comparisons are reported in Supplementary Table S9; further specifications appear in Supplementary Methods S2.

Within each training fold, missing values were median-imputed, fixed missingness indicators were retained for all original features, and imputed features were standardized. Missingness indicators were entered separately, yielding 36, 38, and 40 encoded columns for B0, B1, and B2. Only transformations learned from training data were applied to validation records. Three-state probabilities were estimated by multinomial logistic regression with an L2 penalty, without class weighting or oversampling [26]. A single multiclass temperature parameter, constrained to 0.5–2, was fitted to pooled nested outer-fold predictions after inner-fold tuning and then applied to validation probabilities. No recalibration was performed in the validation set. Primary AUCs used raw probabilities, whereas prediction-error and net-benefit analyses used development-temperature-calibrated probabilities; results for both probability versions were retained.

For the new extensions, the ridge family and common C values of 0.03 and 0.003 were fixed without further tuning. All five blocks (B1, B2, G, S, and SC) used identical community folds. Alongside raw and temperature-calibrated probabilities, symmetric intercept updates estimated two offsets relative to state 0 and renormalized the three probabilities to sum to one. Offsets and temperatures were alternative updates fitted only to cross-fitted development predictions. Final common-C models used five-fold development predictions for calibration; the original-model intercept diagnostic used the original nested predictions. For complete outer-fold evaluation of B1/B2 and S/SC, calibrators were fitted using four-fold predictions from each outer training set. The fitted calibrators were then applied to the corresponding outer evaluation fold. Probability updates used development labels exclusively.

2.6. Performance evaluation and additional analyses

The primary comparison was the difference in raw care-need AUC between B2 and B1. Secondary measures were binary care-need Brier score, total three-state Brier score, multiclass log loss, calibration, and net benefit. The binary Brier score was the mean squared error for state 1; the three-state score summed squared errors across states without dividing by three. Calibration was assessed using mean predicted minus observed probability, a joint logistic calibration intercept and slope, and plots of five approximately equal-sized probability groups within each model [17,18]. Calibration estimates described validation performance and were not used to update predictions. Net benefit at illustrative thresholds of 5% and 10% was $\frac{TP}{n} - \frac{FP}{n} \times \frac{t}{1-t}$, with all three states included in n [19]. TP and FP denote true- and false-positive counts, respectively, and t denotes the assessment threshold. A predicted care-need probability greater than or equal to the threshold defined a positive assessment. These illustrative thresholds represented selection for further functional or care assessment rather than guideline-endorsed cutoffs [27]. Net benefit was expressed as net true-positive equivalents per 1,000 participants, not as an observed intervention effect.

We used 1,000 paired community bootstrap samples. Each resample drew entire communities and used the same records for both models. We reported percentile 95% confidence intervals (CIs) and approximate two-sided Wald P values. The Wald statistic was the observed difference divided by its bootstrap standard deviation. Fitted models, temperatures, and observation weights remained fixed. The intervals therefore quantified uncertainty conditional on the validation sample and did not encompass the full development and selection process. The original exploration considered care need and cardiovascular outcomes together. The Holm adjustment for its two primary AUC tests was retained in the care-need analysis.

The common-parameter, interview-assistance, and observation-weighting analyses were specified after the primary results were known, with their settings fixed before those additional analyses. B1 and B2 were compared at common C values of 0.03 and 0.003, with five community folds for out-of-fold temperature estimation followed by full-development fitting. Both raw and calibrated Brier scores were examined, and the two AUC comparisons formed a Holm-adjustment family. Interview-assistance restrictions were applied to survivors while retaining deaths (S2R). After the original interview-assistance measurements were checked, a further restriction was specified using the 2020 health-module field db045 to retain survivors who were assisted never or some of the time; the separate proxy indicator was missing for all S2R validation survivors. This low-

assistance analysis (N2V) used the S2R models without refitting. Detailed wave-specific rules and the timing of these decisions are provided in Supplementary Methods S3.

In the primary validation risk set of 2,016 participants, outcome ascertainment was modeled from prediction-start B1 variables using binary ridge logistic regression with C=1 and 5-fold community cross-fitting. Stabilized inverse-probability weights were applied to the 1,901 participants with known outcomes. Observation probabilities had a lower bound of 0.05, and weights were truncated at the 1st and 99th percentiles; the truncation proportion and effective sample size were reported. Weighting addressed unknown outcomes only within the CysC-measured risk set and depended on conditional ignorability of missingness and positivity. It did not correct selection into blood collection or CysC testing, nor did it recover unobserved care needs before death. Interview-assistance restrictions, observation-weighted results, and other secondary P values were interpreted as exploratory.

Additional analyses comprised development using only the 2013–2015 window (retaining the main-analysis model family and block-specific C values), complete-case validation, and evaluation in earlier holdout windows. Models were not estimated for analyses with insufficient outcome classes; the two-wave change extension also required at least 50 validation care-need states. Modeling used Python 3.12.14 and scikit-learn 1.9.1; figures were prepared in R 4.6.0. Model specifications, source-variable mappings, and additional performance results are described in Additional file 1. Precision–recall performance was summarized by average precision. Prediction distributions and paired probability shifts were displayed descriptively. Raw binary Brier differences were partitioned by observed state using the full validation sample as the denominator. Net-benefit increments across thresholds of 1–15% were explored using 1,000 additional whole-community bootstrap samples with fixed predictions; these intervals were pointwise. Graphical summaries display original sample estimates alongside the interquartile ranges, medians, and percentile 95% intervals of paired community bootstrap differences. These summaries are conditional on the fitted models.

The extension specifications, including S1 at both common C values, were frozen on 11 October 2026 after the original results were known. S1 used only 2013–2015 data for development and evaluated the models in the unchanged later sample. At fixed assessment capacities of 10%, 20%, and 30%, each model selected the highest-risk floor(capacity × sample size) records; ties were resolved by fixed participant identifier and window order. Every paired community bootstrap resample repeated this ranking and capacity rule. We reported captured and missed care-need states per 1,000, sensitivity, positive predictive value, and deaths among those

flagged. Capacities were hypothetical scenarios for in-depth comprehensive assessment, not standards for routine nutritional screening. All planned conditions were retained. New intervals used 1,000 paired community resamples with fixed models and calibrators; intervals for outer-fold predictions likewise did not include repeated model development or earlier research selection. Full specifications are in Supplementary Methods S10.

The authors used OpenAI GPT-6 Astra tools to assist with English language editing. The authors reviewed the edited text and take responsibility for the final manuscript.

3. Results

3.1. Participant characteristics

After baseline eligibility assessment, the development risk sets contained 2,352 and 3,815 person-windows in 2013–2015 and 2015–2018. Excluding 115 and 198 windows with unknown outcomes left 5,854 person-windows from 4,385 participants in 290 communities; 1,469 participants contributed both windows. There were 5,223 states without assistance needs, 351 care-need states, and 280 deaths. The primary validation risk set contained 2,016 participants; 115 had unknown outcomes, leaving 1,901 participants in 150 communities. These included 1,734 without assistance needs, 98 with care needs (5.16%), and 69 deaths (3.63%) (Figure 1; Table 1). Earlier-wave care history was unknown at least once for 635 development person-windows and 349 validation participants; eligibility therefore established the absence of care needs at the required baseline assessments, with earlier history incompletely observed.

Table 1. Participant characteristics across the three main analysis windows

Characteristic	2013→2015 development	2015→2018 development	2018→2020 validation
Analyzed participants/person-windows	2237	3617	1901
Care-need status, n (%)	137 (6.1)	214 (5.9)	98 (5.2)
Death, n (%)	84 (3.8)	196 (5.4)	69 (3.6)
Age (years)	67.00 (63.00, 73.00)	66.00 (63.00, 72.00)	67.00 (64.00, 72.00)
Female	1105 (49.4)	1789 (49.5)	929 (48.9)
Primary education	490 (24.3)	759 (23.7)	424 (22.3)
Junior secondary education or above	369 (18.3)	695 (21.7)	459 (24.1)
Hypertension	775 (34.9)	1294 (36.6)	854 (44.9)
Diabetes	183 (8.3)	338 (9.4)	254 (13.4)
Number of BADL difficulties	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)
Difficulty walking 100 m	119 (5.3)	236 (6.5)	105 (5.5)
CES-D 10 score	7.00 (4.00, 11.00)	6.00 (3.00, 11.00)	7.00 (3.00, 12.00)
Fair or worse self-rated health	1755 (78.6)	2808 (78.8)	1449 (77.4)
Number of IADL difficulties	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)	0.00 (0.00, 0.00)

Characteristic	2013→2015 development	2015→2018 development	2018→2020 validation
Number of mobility difficulties	1.00 (0.00, 2.00)	1.00 (0.00, 2.00)	1.00 (0.00, 2.00)
Creatinine (mg/dL)	0.78 (0.68, 0.92)	0.79 (0.67, 0.92)	0.79 (0.68, 0.93)
Most recent body mass index (kg/m²)	22.70 (20.42, 25.27)	22.99 (20.66, 25.41)	23.45 (21.15, 25.95)
Current smoking at blood wave	686 (31.0)	1054 (29.2)	547 (28.8)
Non-HDL cholesterol (mg/dL)	139.18 (115.59, 165.46)	130.89 (109.65, 152.90)	131.66 (110.42, 154.44)
Glycated hemoglobin (%)	5.20 (4.90, 5.50)	5.80 (5.60, 6.20)	5.80 (5.60, 6.20)
Most recent systolic blood pressure (mmHg)	131.50 (118.50, 146.50)	129.00 (116.50, 143.50)	128.50 (115.50, 141.50)
Cystatin C (mg/L)	1.03 (0.90, 1.19)	0.88 (0.77, 0.99)	0.88 (0.76, 0.98)
Grip strength (kg), exploratory	29.00 (23.00, 36.00)	29.00 (23.00, 36.00)	29.90 (24.00, 37.00)
C-reactive protein (mg/L), exploratory	1.09 (0.60, 2.28)	1.40 (0.80, 2.60)	1.40 (0.80, 2.70)
Hemoglobin (g/dL), exploratory	14.20 (12.90, 15.40)	13.50 (12.40, 14.60)	13.70 (12.60, 14.90)
Nominal visit interval (years)	2.00 (2.00, 2.00)	3.00 (3.00, 3.00)	2.00 (2.00, 2.00)

Note. Continuous variables are summarized as medians (25th, 75th percentiles); binary variables are counts (%), with percentages based on nonmissing observations for each variable. Missing counts are provided in Table S1. Exploratory grip strength, CRP, and hemoglobin are measured clinical correlates and do not define malnutrition or sarcopenia. Development columns count person-windows and cannot be added to obtain independent participant counts. CysC and creatinine are summarized on their original, untransformed scales. Abbreviations: BADL, basic activities of daily living; IADL, instrumental activities of daily living; CES-D 10, 10-item Center for Epidemiologic Studies Depression Scale; HDL, high-density lipoprotein; CysC, cystatin C. To obtain SI values, multiply creatinine in mg/dL by 88.4 for μmol/L and non-HDL cholesterol in mg/dL by 0.02586 for mmol/L.

Median age was 67, 66, and 67 years across the two development windows and primary validation, and women accounted for 49.4%, 49.5%, and 48.9% of participants, respectively. Median CysC concentrations were 1.03, 0.88, and 0.88 mg/L, and median systolic pressures were 131.5, 129.0, and 128.5 mmHg (Table 1). Predictor missingness is reported in Table S1.

3.2. Primary validation performance

Multinomial ridge logistic regression was selected in development. Final C values were 0.03, 0.03, and 0.003 for B0, B1, and B2, with temperatures of 1.0015, 1.0012, and 0.9902. Raw AUCs were 0.7174 (95% CI 0.6510 to 0.7807), 0.7205 (0.6547 to 0.7821), and 0.7214 (0.6518 to 0.7862), respectively (Table 2; Figure 2a). The B2−B1 AUC difference was 0.0009 (−0.0096 to 0.0126), with P=0.878; the Holm-adjusted P value for the original two-outcome primary comparison was 1.000. Conditional AUCs among survivors were 0.7281 for B1 and 0.7296 for B2. Average precision was 0.1476 for B0, 0.1498 for B1, and 0.1588 for B2 (Figure 2b). Raw care-need probabilities by observed next-visit state are shown in Figure 2c.

Table 2. Model performance in primary validation, with 95% CIs

Performance measure	B0	B1	B2
Raw care-need AUC	0.7174 (0.6510 to 0.7807)	0.7205 (0.6547 to 0.7821)	0.7214 (0.6518 to 0.7862)
Calibrated binary Brier	0.048151 (0.040467 to 0.056646)	0.047950 (0.040270 to 0.056474)	0.046605 (0.038852 to 0.055257)
Calibrated three-state Brier	0.154909 (0.134537 to 0.174972)	0.154380 (0.133804 to 0.174625)	0.152692 (0.131918 to 0.173439)

Performance measure	B0	B1	B2
Multiclass log loss	0.323940 (0.287992 to 0.358806)	0.322752 (0.286456 to 0.358054)	0.319265 (0.282889 to 0.355647)
Calibrated mean predicted−observed difference	0.02264 (0.01240 to 0.03228)	0.02201 (0.01180 to 0.03161)	0.00863 (−0.00121 to 0.01792)
Net benefit at 5% threshold	0.01337 (0.00427 to 0.02314)	0.01415 (0.00497 to 0.02425)	0.01717 (0.00839 to 0.02663)
Net benefit at 10% threshold	0.00812 (0.00058 to 0.01630)	0.00725 (0.00006 to 0.01485)	0.00994 (0.00303 to 0.01686)

Note. Validation n=1,901, including 98 care-need states and 69 deaths. B0 includes creatinine and other laboratory variables; B1 adds systolic blood pressure; B2 adds CysC. Intervals are conditional estimates from 1,000 community bootstrap samples. AUC uses raw probabilities; other measures use development-temperature-calibrated probabilities. Net benefit is per participant; multiplication by 1,000 gives net true-positive equivalents per 1,000 participants, as shown in the text and figures. Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; CysC, cystatin C.

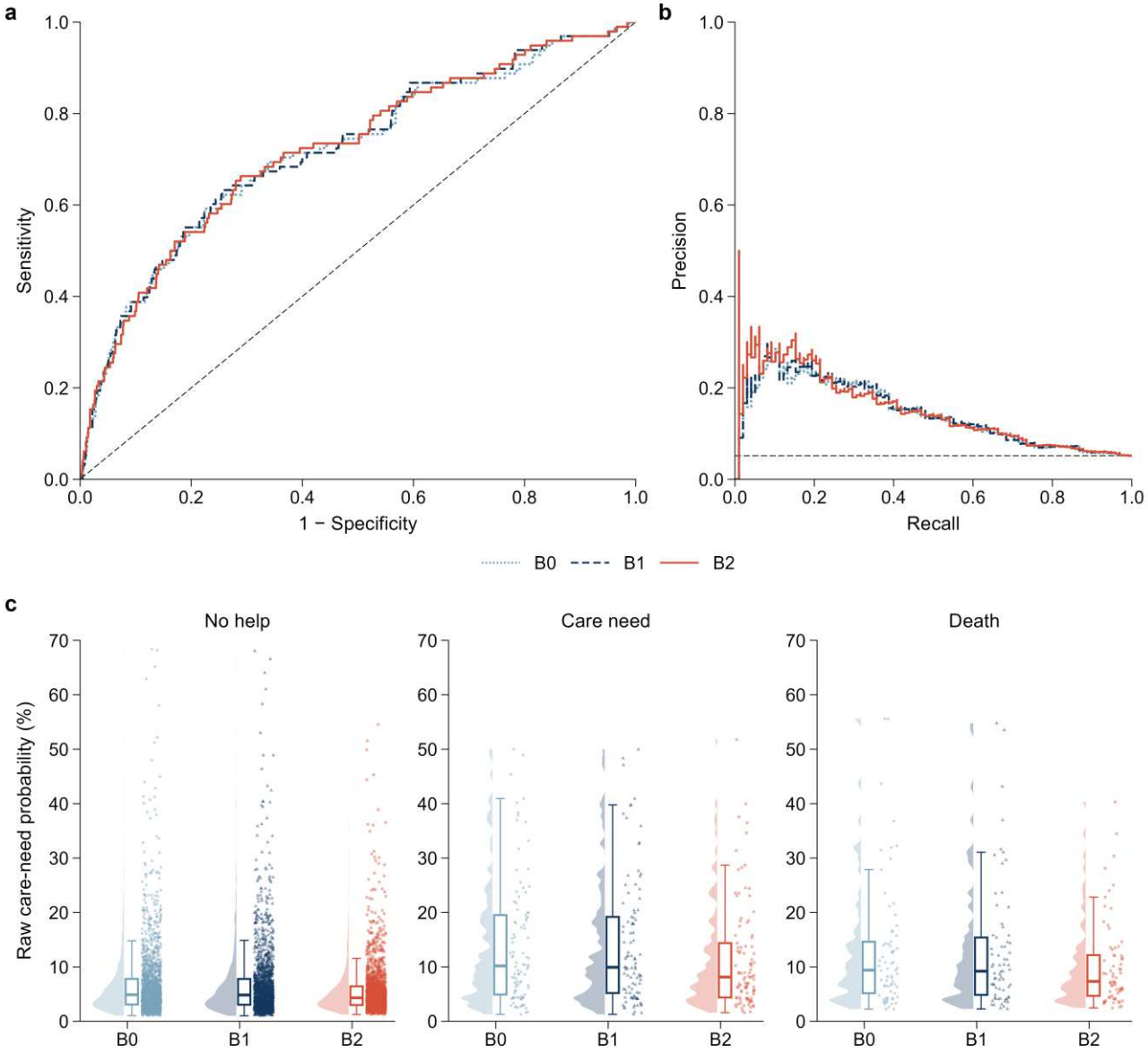

Figure 2. Care-need discrimination with and without cystatin C. B0 includes clinical, functional, and laboratory predictors, including creatinine; B1 adds the most recently available systolic blood pressure; B2 adds cystatin C. a, Empirical ROC curves use fixed raw care-need probabilities. AUC (95% community-bootstrap CI): B0 0.717 (0.651–0.781); B1 0.720 (0.655–0.782); B2 0.721 (0.652–0.786). The diagonal denotes chance-level discrimination. b, Empirical precision–recall curves. The horizontal reference denotes care prevalence (98/1,901; 5.16%). Average precision: B0 0.148, B1 0.150, B2 0.159. c, Raw risks by observed state: no help (n=1,734), care need (n=98), and death (n=69). Each group includes all observations for each model. Group-normalized density widths do not encode sample size. Boxes show median and quartiles, with Tukey 1.5-IQR whiskers. Points represent individual predictions, including observations beyond the whiskers. All three state panels display the full risk range. Death and no help are negative states for the binary ROC/PR target. Precision–recall curves, average precision, and distributions are descriptive. ROC, receiver operating characteristic; AUC, area under the ROC curve; CI, confidence interval; PR, precision–recall; IQR, interquartile range.

Calibrated binary care-need Brier scores were 0.047950 for B1 and 0.046605 for B2, a difference of −0.001345 (95% CI −0.002304 to −0.000288), or an approximately 2.8% relative reduction. The corresponding raw binary Brier difference was −0.001318 (95% CI −0.002265 to −0.000285). The total three-state Brier difference was −0.001688 (−0.003740 to 0.000528), and the multiclass log-loss difference was −0.003487 (−0.007562 to 0.000727). Binary care-need error was lower, whereas the reduction in overall three-state error remained uncertain (Tables 2 and 3).

Table 3. Incremental value of CysC in the main and common-parameter comparisons

Analysis	ΔAUC (95% CI)	Raw ΔBrier (95% CI)	Calibrated ΔBrier (95% CI)	ΔNet benefit at 5% (95% CI)
Separately tuned main strategy	0.0009 (−0.0096 to 0.0126)	−0.001318 (−0.002265 to −0.000285)	−0.001345 (−0.002304 to −0.000288)	0.00302 (0.00083 to 0.00484)
Common C=0.03	0.0016 (−0.0066 to 0.0097)	−0.000554 (−0.000897 to −0.000182)	−0.000570 (−0.000910 to −0.000202)	0.00053 (−0.00239 to 0.00306)
Common C=0.003	0.0005 (−0.0078 to 0.0086)	−0.000235 (−0.000457 to 0.000033)	−0.000213 (−0.000440 to 0.000057)	0.00125 (−0.00068 to 0.00281)

Note. All differences are B2−B1. Positive AUC/net-benefit differences and negative Brier differences favor B2. Both common-parameter AUC comparisons had Holm-adjusted P=1.000; other measures are secondary and exploratory. Raw Brier comparisons at common C exclude differences in temperature parameters. Net benefit is per participant; multiplication by 1,000 gives net true-positive equivalents per 1,000 participants, as shown in the text and figures. Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; CysC, cystatin C. C denotes inverse regularization strength.

The observed care-need proportion was 5.16%, compared with mean calibrated predicted probabilities of 7.36% for B1 and 6.02% for B2. The joint calibration intercept and slope were −0.661 and 0.886 for B1, and −0.106 (95% CI −0.850 to 0.660) and 1.027 (0.734 to 1.343) for B2. The B2 point estimates were closer to ideal calibration, although the intervals remained wide. Development temperatures were close to 1. The original temperature-calibrated and new intercept-updated curves were similar (Figure 3a,b); applying calibration learned in development did not eliminate overprediction in the later sample. The model-specific quintiles are descriptive; the original joint calibration estimates and their marginal intervals are retained in Table S2.

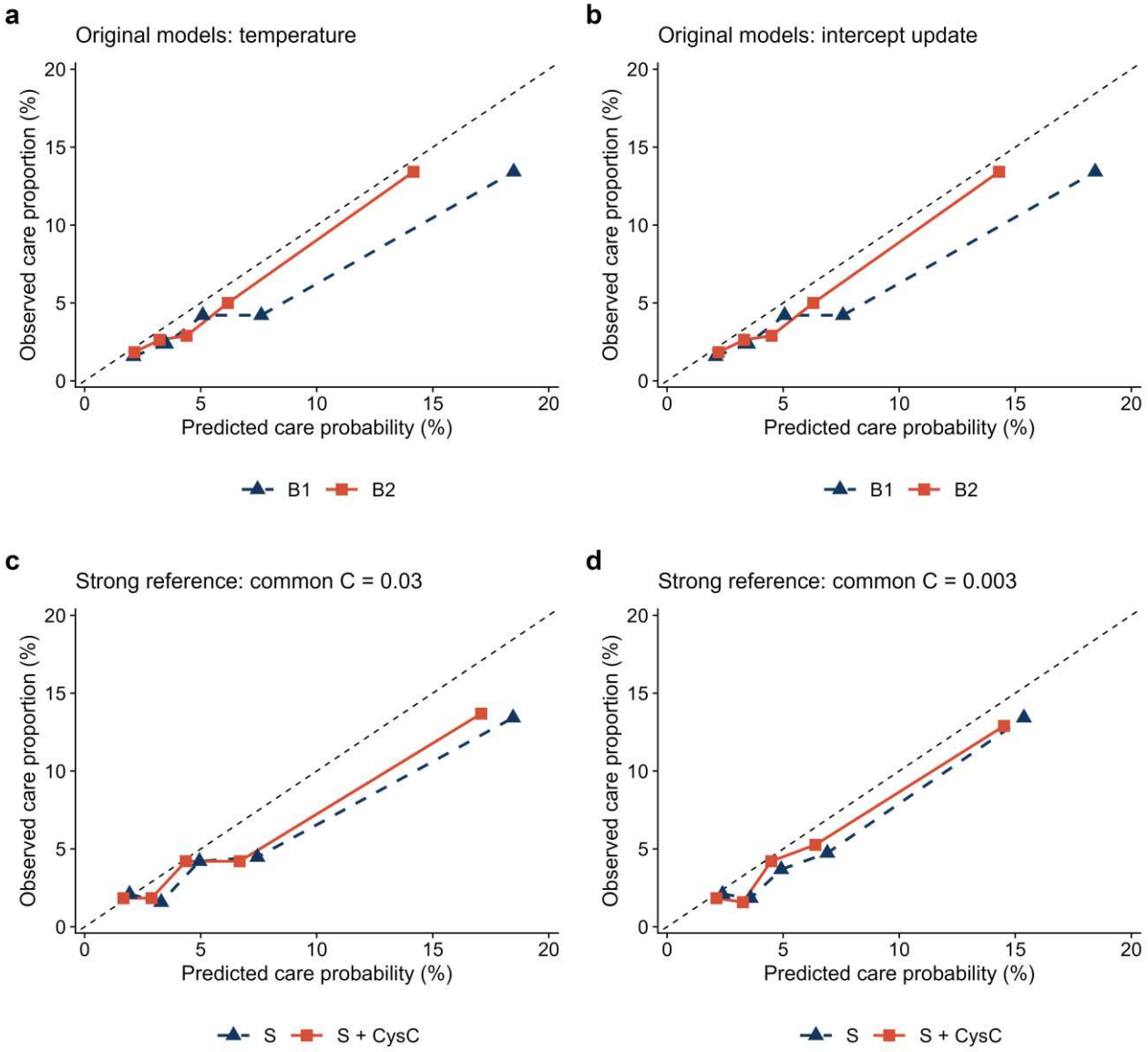

Figure 3. Care-need calibration under symmetric development-only updating. All panels include the same 1,901 participants, with 98 care-need states and 69 deaths. a, Original separately tuned B1/B2 models with their original temperature calibration. b, The same models with two intercept offsets estimated from their original nested development predictions. c,d, Enhanced clinical reference S and S plus CysC at common C=0.03 and C=0.003, respectively, with two intercept offsets estimated from five-fold development predictions. S comprises B1 plus continuous grip strength, natural-log CRP, and hemoglobin. Offsets use no-help as the reference state and renormalize all three probabilities; no evaluation-set updating was performed. Points show mean predicted probability and observed care proportion in five model-specific groups of 380–381 participants. Lines connect the points, and diagonals denote ideal calibration. Quintiles may contain different participants across models; inferential intervals are not shown. Death is non-care in this binary display. Blue triangles/dashed lines identify the reference and vermilion squares/solid lines identify its CysC extension. Raw probabilities, temperature alternatives, and state-specific diagnostics are reported in Tables S14 and S19. C, inverse regularization strength; CRP, C-reactive protein.

At the 5% threshold, net benefit was 14.15 for B1 and 17.17 for B2 in net true-positive equivalents per 1,000 participants, an increment of 3.02 (95% CI 0.83 to 4.84). At the 10% threshold, the increment was 2.69 per 1,000 (−0.58 to 5.77). The full-model point estimates exceeded both assess-all and assess-none strategies at both thresholds, but the improvement of B2 over B1 varied by threshold (Table 2; Figure 4a). Assess-all net benefit was 1.63 per 1,000 at 5% and −53.83 at 10%; assess-none net benefit was zero at both thresholds. At

5%, B2 classified 800 participants for further assessment compared with 969 for B1; these included 71 and 74 observed care-need states, respectively. Thus, B2 flagged 169 fewer participants, comprising 166 fewer false positives and three fewer true positives. At 10%, the number flagged decreased from 356 to 230 and the number of true positives from 48 to 40: 126 fewer flags and eight fewer true positives. These changes show the trade-off between fewer assessment flags and fewer identified care-need states (Figure 4b).

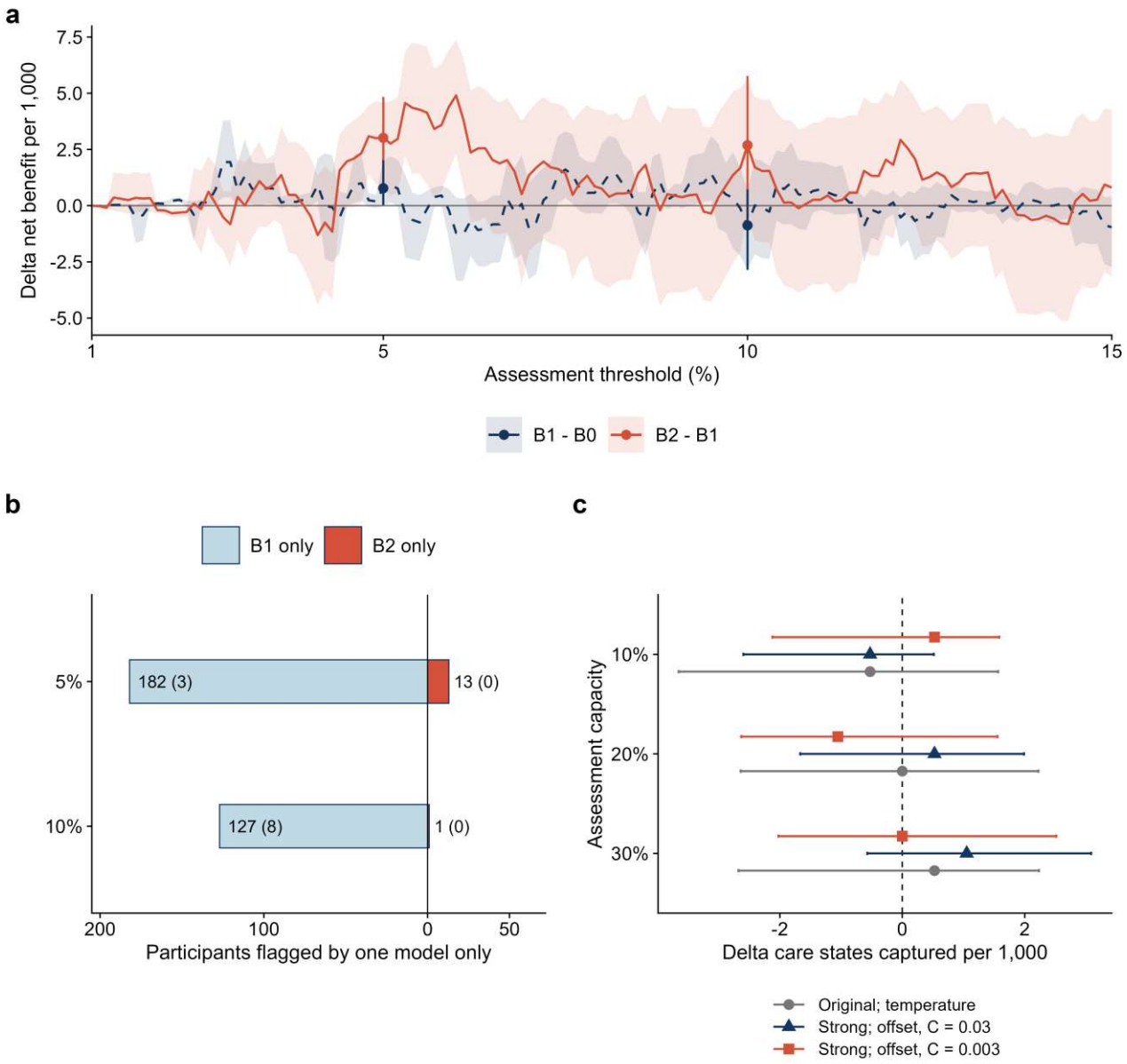

Figure 4. Assessment trade-offs at probability thresholds and equal capacities. B1 contains clinical, functional, and laboratory predictors including systolic blood pressure; B2 adds CysC. a, Original development-temperature-calibrated incremental net benefit across 1%–15% thresholds. Net benefit is $\frac{TP-FP\times\frac{t}{1-t}}{1{,}901}$, displayed per 1,000. Shading shows pointwise 95% percentile intervals from 1,000 whole-community bootstrap resamples with fixed models and temperatures; these are not simultaneous bands. Points and vertical intervals at 5% and 10% retain the original paired estimates and intervals. b, Original paired classifications at 5% and 10%. Leftward/rightward bars show B1-only/B2-only flags; their direction indicates category, not negative counts. Labels show total n (care-need n). At 5%, both models were below threshold for 919 (24) and above threshold for 787 (71); at 10%, corresponding counts were 1,544 (50) and 229 (40). c, CysC-minus-reference differences in care states captured per 1,000 at 10%, 20%, and 30% assessment capacities. Gray circles show original B2−B1 with temperature calibration; blue triangles and vermilion squares show SC−S with intercept updating at common C=0.03 and C=0.003. S comprises B1 plus grip strength, natural-log CRP, and hemoglobin; SC adds CysC. The capacity rule selected 190, 380, or 570 of the same 1,901 participants. Each of 1,000 paired community resamples recomputed the sample size, capacity, and risk ranking under the frozen tie rule; bars are conditional percentile 95% intervals. Higher values favor CysC. Death remains a non-care state; deaths among flags and all other capacity outcomes are reported in Tables S16 and S17. Capacities are hypothetical scenarios for in-depth assessment, not eligibility rules for routine nutritional screening or observed clinical benefit. TP, true positives; FP, false positives; t, probability threshold.

3.3. Incremental value at common regularization parameters

The raw B2−B1 AUC difference was 0.0016 (−0.0066 to 0.0097) at C=0.03 and 0.0005 (−0.0078 to 0.0086) at C=0.003; both Holm-adjusted P values were 1.000. At C=0.03, the calibrated binary Brier difference was −0.000570 (−0.000910 to −0.000202), and the raw difference was −0.000554 (−0.000897 to −0.000182). At C=0.003, the corresponding differences were −0.000213 (−0.000440 to 0.000057) and −0.000235 (−0.000457 to 0.000033). A reduction in binary prediction error was therefore observed before and after calibration under weaker shrinkage, whereas the corresponding intervals crossed zero under stronger shrinkage (Table 3; Figure 5a).

At the 5% threshold, net-benefit increments at common C=0.03 and C=0.003 were 0.53 (95% CI −2.39 to 3.06) and 1.25 (−0.68 to 2.81) net true-positive equivalents per 1,000 participants, respectively (Table 3; Figure 5a). At 10%, the corresponding increments were 1.17 per 1,000 (−1.64 to 3.58) and −0.88 per 1,000 (−4.02 to 1.79). Net-benefit increments were uncertain at both common C values, and the main contrast also incorporated differences in shrinkage. Across the three regularization strategies, paired raw probability changes were predominantly negative (Figure 5b). In the main comparison, the contributions to the raw binary Brier difference, multiplied by 1,000, were −3.03, +2.18, and −0.47 for no-help, care-need, and death states, respectively, summing to approximately −1.32. Each contribution used the full validation sample of 1,901 as its denominator (Figure 5c).

For overall three-state performance, total Brier differences at C=0.03 and 0.003 were −0.000787 (−0.002026 to 0.000450) and −0.000561 (−0.001320 to 0.000232), respectively. Multiclass log-loss differences were −0.002775 (−0.006055 to 0.000557) and −0.002784 (−0.004795 to −0.000593). The log-loss interval at C=0.003 excluded zero, whereas gains in AUC and net benefit remained uncertain.

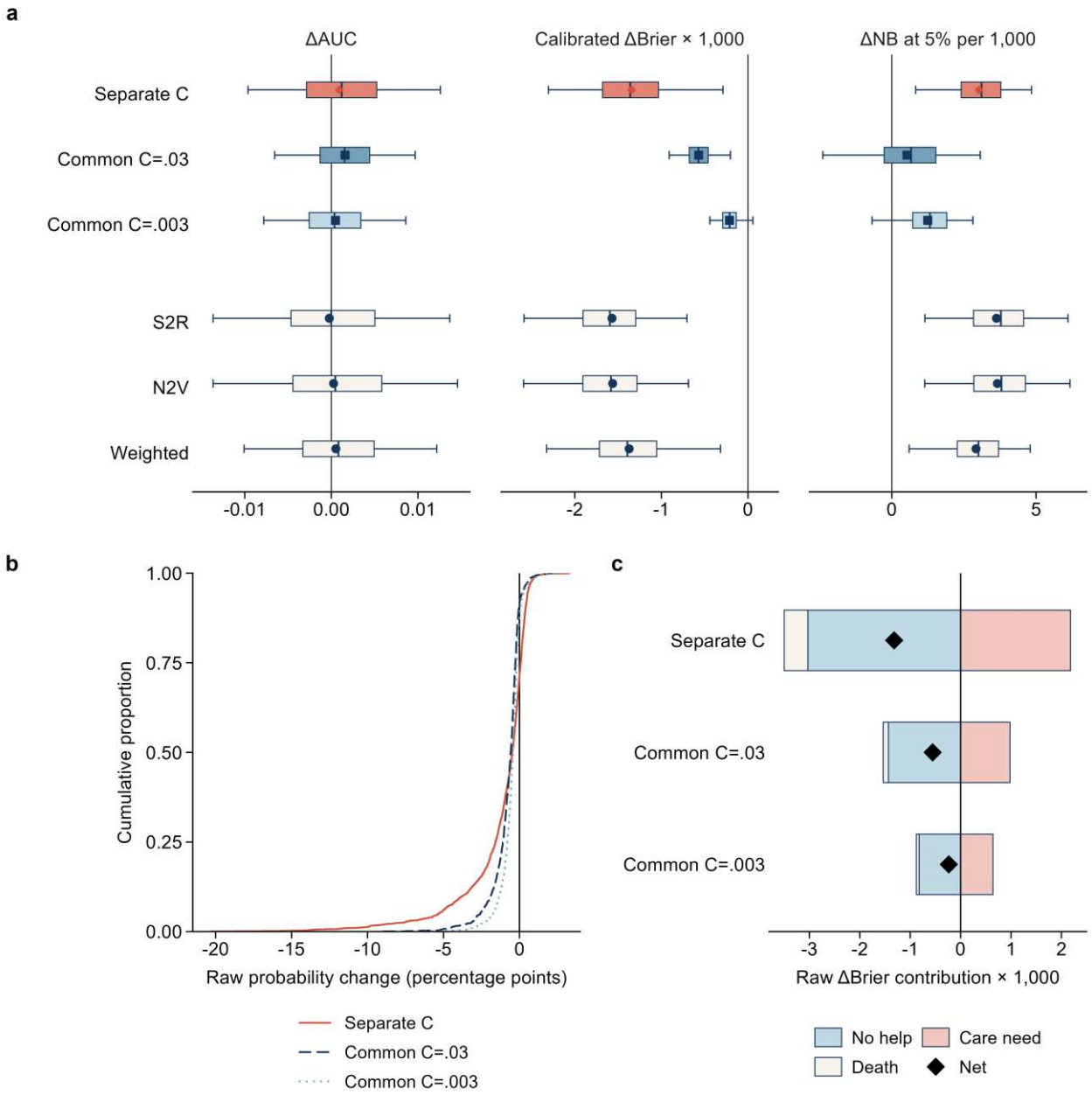

Figure 5. Cystatin C increments varied with regularization strength. B2 adds cystatin C to B1. a, Paired B2−B1 differences in raw AUC, development-temperature-calibrated binary Brier (×1,000), and calibrated net benefit at a 5% threshold (per 1,000). Boxes show Q25–Q75, internal lines Q50, and whiskers Q2.5–Q97.5 from 1,000 paired community bootstrap samples, with models, temperatures, and weights fixed. Whiskers are conditional percentile 95% intervals, not Tukey whiskers. Overlaid points indicate estimates in the original sample. Positive ΔAUC/ΔNB and negative ΔBrier favor B2. Vermilion diamonds denote separate C (B1=0.03; B2=0.003); blue squares denote common C=0.03 or 0.003. Blue circles denote interview-assistance restrictions retaining deaths (S2R, n=1,755), additional lower-assistance restriction reusing S2R models (N2V, n=1,726), and observation weighting (Weighted, n=1,901). b, Exact empirical cumulative distributions of raw paired probability changes, $100 \times (p_{B2} - p_{B1})$; negative values indicate lower predicted probability. c, Signed no-help, care-need, and death contributions to raw binary ΔBrier. Each equals the state-specific sum of $\frac{(p_{B2}-Y)^2-(p_{B1}-Y)^2}{1{,}901} \times 1{,}000$, where $Y = I(state = care\ need)$. Black diamonds show the algebraic sum. Panels b and c include 1,901 participants per strategy (1,734 no-help, 98 care-need, 69 deaths). Death is a non-care outcome in this error measure. These panels are descriptive, without additional intervals or causal interpretation. Raw Brier intervals are reported in Table 3; the interval at common C=0.003 includes zero. AUC, area under the receiver operating characteristic curve; NB, net benefit; Q, percentile; C, inverse regularization strength.

3.4. Sensitivity analyses

The S2R interview-assistance restriction yielded 5,308 development person-windows from 4,036 participants and 1,755 validation participants, including 68 care-need states and 61 deaths. The separate 2020 proxy indicator was missing for all 1,694 survivors. Applying the additional db045 restriction removed 29 survivors who were assisted most or all of the time and retained 1,726 participants, including 66 care-need states and all 61 deaths (N2V). The B2−B1 AUC difference was 0.0003 (95% CI −0.0137 to 0.0146), and the calibrated binary Brier difference was −0.001562 (−0.002591 to −0.000687) (Table 4; Figure 5). These results apply to participants with low interview assistance; proxy-response status remained uncertain.

Table 4. Additional sensitivity analyses

Analysis	Participants/care need/death	B2 AUC (95% CI)	ΔAUC (95% CI)	Calibrated ΔBrier (95% CI)
Restricted interview assistance (S2R)	1755/68/61	0.7101 (0.6306 to 0.7901)	−0.0002 (−0.0137 to 0.0137)	−0.001570 (−0.002586 to −0.000703)
Low assistance at validation (N2V)	1726/66/61	0.7096 (0.6285 to 0.7924)	0.0003 (−0.0137 to 0.0146)	−0.001562 (−0.002591 to −0.000687)
Outcome-observation weighting	1901/98/69	0.7209 (0.6510 to 0.7850)	0.0005 (−0.0101 to 0.0122)	−0.001372 (−0.002325 to −0.000317)

Note. S2R restricts interview assistance in development and validation while retaining deaths; N2V applies the additional lower-assistance restriction to validation participants using the S2R models. The S2R 2020 proxy indicator was missing for all survivors. N2V restricts interview assistance using the original db045 field, but does not establish strictly independent participant responses. The observation-weighted bootstrap holds estimated weights fixed; effective n=1,899.1. Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval. Differences are B2 minus B1; B1 includes clinical, functional, and laboratory predictors with systolic blood pressure, and B2 adds cystatin C.

Outcome-observation weighting yielded an effective sample size of 1,899.1. The weighted B2−B1 AUC difference was 0.0005 (95% CI −0.0101 to 0.0122), the calibrated binary Brier difference was −0.001372 (−0.002325 to −0.000317), and the net-benefit increment at 5% was 2.92 net true-positive equivalents per 1,000 participants (0.60 to 4.80). Weight diagnostics are provided in Supplementary Methods S4, and selection comparisons are shown in Tables S11 and S12. These results were similar to those of the main analysis within the CysC-measured risk set; weighting addressed outcome observation, whereas selection into blood measurement remained outside its scope.

When development was restricted to 2013–2015, the AUC increment was 0.0095 (95% CI −0.0041 to 0.0240) and the calibrated binary Brier difference was −0.001770 (−0.003168 to −0.000529). Calibrated net-benefit increments were 0.91 (−1.37 to 2.81) and 2.63 (−0.44 to 5.27) net true-positive equivalents per 1,000 at 5% and 10%, respectively (Table S10). Complete-case validation gave an AUC increment of 0.0052 (−0.0074 to 0.0182). Earlier holdout windows did not show a consistent CysC advantage (Tables S3 and S4). The original low-proxy branch was not estimable because it excluded all development deaths, and the two-wave change extension had 874 validation participants and 47 care-need states, below the analysis threshold of 50.

3.5. Exploratory clinical, calibration, and capacity extensions

After quality control, grip strength was available for 1,842/2,237, 3,548/3,617, and 1,874/1,901 records across the three windows; hemoglobin was available for 2,179, 3,566, and 1,887, and CRP for all records. Missingness did not change the comparison samples. Adding grip strength to B1 yielded raw AUC increments of 0.0051 at each C, with intervals crossing zero; adding CRP and hemoglobin also produced uncertain discrimination increments (Tables S13–S15).

For S versus SC, raw AUCs were 0.7258/0.7265 at C=0.03 and 0.7294/0.7282 at C=0.003. The respective increments were 0.0007 (95% CI −0.0066 to 0.0072) and −0.0012 (−0.0088 to 0.0052). Intercept-updated binary Brier differences were −0.000459 (−0.000760 to −0.000132) and −0.000201 (−0.000424 to 0.000038), respectively (Table 5; Figure 3c,d). Symmetric intercept updating preserved the regularization-dependent pattern for B1/B2 as well. All common-C B1/B2 and S/SC net-benefit intervals at 5% and 10% crossed zero under intercept updating (Tables S14 and S15).

Table 5. Exploratory CysC increments under common regularization and symmetric calibration

Contrast	C	Raw ΔAUC (95% CI)	Raw ΔBrier (95% CI)	Temperature ΔBrier (95% CI)	Intercept-update ΔBrier (95% CI)
B2−B1	0.03	0.0016 (−0.0067 to 0.0094)	−0.000554 (−0.000914 to −0.000183)	−0.000570 (−0.000927 to −0.000200)	−0.000549 (−0.000904 to −0.000180)
B2−B1	0.003	0.0005 (−0.0084 to 0.0082)	−0.000235 (−0.000489 to 0.000034)	−0.000213 (−0.000470 to 0.000062)	−0.000233 (−0.000483 to 0.000034)
SC−S	0.03	0.0007 (−0.0066 to 0.0072)	−0.000462 (−0.000771 to −0.000130)	−0.000477 (−0.000782 to −0.000149)	−0.000459 (−0.000760 to −0.000132)
SC−S	0.003	−0.0012 (−0.0088 to 0.0052)	−0.000203 (−0.000429 to 0.000042)	−0.000181 (−0.000409 to 0.000062)	−0.000201 (−0.000424 to 0.000038)

Note. B2=B1+CysC; S=B1+grip strength+ln(CRP)+hemoglobin; SC=S+CysC. Brier denotes the binary care-need Brier score. Positive AUC and negative Brier differences favor the added CysC information. All models use the same 1,901 participants. New intervals use 1,000 paired community bootstrap samples with fixed models/calibrators; their random draws differ from the previously reported intervals in Table 3, which are retained. Temperature and intercept updates are separate development-only alternatives. C is inverse regularization strength. These extensions, specified after the original results were known, are exploratory; complete performance, all sequence steps, and net-benefit estimates are in Tables S14 and S15.

At capacities of 10%, 20%, and 30%, the intercept-updated S/SC models captured 35/34, 51/52, and 60/62 care states at C=0.03, and 33/34, 51/49, and 62/62 at C=0.003. All paired care-capture intervals crossed zero (Figure 4c; Table S16). S1 analyses at common C values gave raw AUC increments of 0.0004 (−0.0045 to 0.0051) and −0.0013 (−0.0082 to 0.0051), respectively. Both S1 intercept-updated Brier intervals were below zero; a positive net-benefit increment at the 10% threshold for C=0.03 was not preserved at C=0.003 (Table S17). In

458 complete outer-fold development evaluation, both S/SC AUC and Brier difference intervals crossed zero at both
459 C values (Table S18). State-specific diagnostics are provided in Table S19.

460 ## 4. Discussion

461 ### 4.1. Principal findings

462 Adding CysC to an established set of clinical, functional, and laboratory predictors produced a small reduction
463 in binary prediction error, with little change in discrimination for next-visit care-need status. The reference and
464 expanded models both had AUCs of approximately 0.72, and the primary increment was 0.0009 (95% CI
465 −0.0096 to 0.0126). The reduction in binary prediction error varied with regularization. The same pattern
466 persisted after accounting for grip strength, CRP, and hemoglobin: increments in discrimination and care
467 capture remained uncertain across all preselected capacities.

468 Lower Brier scores and nearly unchanged AUCs reflect different aspects of prediction: AUC measures ranking,
469 whereas the Brier score measures probability error [16,17]. The reference model overestimated the average
470 probability of care need, and the expanded model reduced this overestimation. Its calibration point estimates
471 were also closer to their ideal values, although the confidence intervals remained wide. This pattern is
472 consistent with improved probability estimation and little change in the model's ability to distinguish people by
473 subsequent assistance needs. Calibration remains relevant to using such probabilities in practice even when
474 discrimination changes little [18].

475 The common-parameter analyses show how the CysC increment depends on shrinkage. Because B1 and B2
476 were tuned separately, the main contrast incorporated both the added CysC information and differences in
477 shrinkage. Estimated penalty parameters can affect calibration and prediction error, making performance
478 conditional on the regularization strategy [28]. At a common C of 0.03, binary Brier scores improved with
479 CysC before and after temperature calibration, supporting an error reduction within that specification. At
480 C=0.003, which imposed stronger shrinkage, the binary Brier interval included zero. The secondary multiclass
481 log-loss result at this setting indicates that any gain also depends on the performance measure. Neither
482 common-parameter analysis established a clear net-benefit increment. This dependence persisted with
483 symmetric intercept updates, and CysC increments beyond the enhanced reference remained uncertain in full

outer-fold evaluation. Using a common C provides a diagnostic comparison, although differences in model complexity remain. Calibration learned in development may also leave shifts in later periods uncorrected.

4.2. Relation to previous evidence

Earlier studies linked CysC-related measures to frailty, physical function, and disability. Li et al. reported associations of CysC-related kidney function measures with frailty progression and decline in physical function in CHARLS and the Health and Retirement Study, including associations that remained after adjustment for creatinine in CHARLS [11]. Bai et al. examined creatinine- and CysC-based indices in relation to incident frailty and ADL disability [12]. Wu et al. reported an association between the creatinine-to-CysC ratio and ADL disability trajectories in 4,617 CHARLS participants [13]. Deng et al. evaluated CysC–creatinine scores for prevalent frailty, while Matsuzawa et al. examined a derived index for sarcopenia incidence or progression [14,15]. Our comparison evaluates the added predictive value for future BADL assistance beyond existing clinical information and measured strength.

Other cohorts have linked CysC-related measures to functional vulnerability. Cross-sectional studies associated differences between CysC- and creatinine-based estimated glomerular filtration rates (eGFRs), and the creatinine-to-CysC ratio, with muscle characteristics or physical performance [29,30]. Prospective studies examined eGFR discordance in relation to falls, hospitalization, and death in US older adults [31], CysC-based kidney measures in relation to motoric cognitive risk in CHARLS and the Health and Retirement Study [32], and plasma CysC in relation to long-term-care disability certification among Japanese octogenarians [33]. These exposures and endpoints differ from adding CysC to an established model of next-visit BADL assistance.

Large meta-analyses have also associated lower eGFR, albuminuria, and discordance between creatinine- and CysC-based eGFR with adverse clinical outcomes [34,35]. Work in chronic kidney disease suggests that several measured determinants contribute to eGFR discordance, without fully explaining it [36]. These prognostic associations support the clinical relevance of CysC-related measures; their incremental value for BADL assessment requires direct evaluation.

The paired comparison assessed CysC beyond a model already containing creatinine, recent functional difficulties, disease information, and measured blood pressure. Previous Chinese studies have developed models for disability or functional dependency, and one CHARLS model considered CysC among its candidate predictors [5–7]. The nested models used the same sample, with identical eligibility criteria and outcome

frequencies across information blocks. This comparison addresses the incremental-value question in biomarker evaluation [16]. CHARLS research on multimorbidity trajectories and subsequent disability also addresses a different exposure and follow-up framework [37]. Differences in functional outcomes, eligibility criteria, follow-up intervals, and evaluation designs preclude direct comparisons of AUCs across these studies.

4.3. Implications for community assessment

Using an existing CysC result and ordering a new test have different implications for community assessment. The observed reduction in probability error was small and did not translate into consistent selection gains beyond measured strength and other clinical information at equal assessment capacity. Ordering a new test would require evidence that the predictive gain improves care decisions sufficiently to justify the added cost and burden. At the illustrative 5% threshold, the main net-benefit increment was about three net true-positive equivalents per 1,000 people. This expresses a specified trade-off between true- and false-positive assessments [19]; it does not count additional people detected or episodes of dependency prevented. Net benefit here concerns being alive with BADL assistance needs at the next visit; it does not quantify benefits for mortality prevention, care needs arising between visits, or nutritional assessment itself. The reduction in assessment flags also included fewer identified care-need states, and the gain was uncertain at common regularization parameters. An independent evaluation should specify the assessment action, threshold, and minimum useful increment in advance. It should then test whether adding CysC improves those decisions [38].

Evidence on community interventions suggests that individualized care planning with medication review and follow-up can help sustain independence, with effects depending on the intervention [39]. Functional limitations and receipt of assistance also need to be distinguished: a multinational study documented gaps in assistance among community-dwelling older adults with dementia and functional limitations [40]. For nutrition care, identifying future assistance needs may help plan comprehensive evaluation. Routine nutritional screening and diagnostic nutritional assessment, when indicated, remain necessary [4,41]. Grip strength measures muscle strength, CRP reflects inflammation, and hemoglobin has several determinants; none defines malnutrition or establishes sarcopenia alone. Nutritional status and intervention benefits were not assessed in this study. Evaluating these outcomes would clarify the clinical implications of CysC-guided selection.

4.4. Strengths and limitations

The study evaluated nested information blocks on the same participants, kept communities separate between development and evaluation, and restricted preprocessing and temperature estimation to development data. A later evaluation window also tested performance after changes in calendar period and predictor timing. These design features support within-study comparisons; interpretation also depends on the risk of bias and applicability [42]. The holdout sample had been used in earlier exploration, and the common-parameter and response-mode analyses, together with the clinical, calibration, and capacity extensions, were specified after earlier results were known. The evaluation is therefore exploratory within CHARLS. Community bootstrap intervals were conditional on the fitted models and did not incorporate the full uncertainty from study-question, model, and tuning selection [43]. The new outer-fold evaluation restricted calibration to training folds; the later evaluation sample remained previously used.

The intended application uses historical biomarker measurements. In the primary 2018→2020 evaluation, laboratory values, blood pressure, body mass index, and grip strength came from the 2015 wave, while functional and disease information came from the 2018 interview. The study evaluates historical CysC information added to a more recent clinical assessment. The benefit of contemporaneous CysC testing at every prediction start remains undetermined. Participant-specific collection dates were unavailable in the released blood files, so the order of blood collection and the interview within the 2015 development wave could not be established. This uncertainty applies to all same-wave development measurements, including the 16 person-windows with interviews in 2016. Visit intervals varied between two and three years. Blood collection conditions differed between waves [21,22], and outcome questionnaire routing also changed. These differences may affect transportability across periods; future evaluation should target the intended population, measurement schedule, and setting [44]. Including the nominal interval in the model did not convert the outcome into a standardized incidence estimate.

All information blocks were evaluated in older adults who provided blood and had CysC measured. The value of testing older adults without an existing CysC result requires separate evaluation. Observation weighting produced similar results within this measured-CysC risk set but depended on measured predictors of outcome ascertainment. It did not correct selection into blood collection or CysC measurement, and it does not exclude missingness related to unmeasured functional deterioration. Missing predictor data were handled consistently

during training and evaluation. The adequacy of this approach depends on the missingness process; missingness indicators alone do not establish it [45]. Grip noncompletion included health-related and other reasons; missingness indicators cannot recover unmeasured strength. Some released hemoglobin values were unusually high and were retained under the frozen positive-value rule, so measurement error remains possible. Restrictions on interview assistance addressed response mode, with residual uncertainty about response independence and measurement error.

The endpoint describes being alive and needing help at a survey visit. Transient care needs between visits, recovery before the next interview, and unrecorded assistance needs before death may therefore be missed. Longitudinal studies of disability and recovery distinguish transitions between states from status recorded at a single follow-up visit [46,47]. Modeling death separately preserves the distinction between death and survival without care need, but it cannot recover an unobserved predeath care history. The 98 care-need states in the primary evaluation limited the precision of calibration and incremental-value estimates. A future validation study should be sized for the desired precision of calibration, discrimination, and decision net benefit in its intended setting [48]. Secondary findings also arose from an exploratory analysis history without unified multiplicity control. The evidence therefore does not establish equivalence of the two models or support a general recommendation for additional CysC testing.

5. Conclusions

In older CHARLS participants with measured CysC, adding historical CysC to clinical, functional, and laboratory information produced a small, regularization-dependent reduction in care-need probability error. Beyond grip strength, CRP, and hemoglobin, increments in discrimination and selection for assessment remained uncertain under symmetric calibration and equal-capacity comparisons. Independent evaluation with a predefined assessment pathway is needed to determine whether additional CysC testing changes care decisions [49].

List of abbreviations

ADL, activities of daily living; AUC, area under the receiver operating characteristic curve; BADL, basic activities of daily living; BMI, body mass index; CES-D 10, 10-item Center for Epidemiologic Studies Depression Scale; CHARLS, China Health and Retirement Longitudinal Study; CI, confidence interval; CRP,

C-reactive protein; CysC, cystatin C; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; FP, false positives; GFR, glomerular filtration rate; Hb, hemoglobin; HDL, high-density lipoprotein; IADL, instrumental activities of daily living; IQR, interquartile range; NB, net benefit; non-HDL, non-high-density lipoprotein; OOF, out-of-fold; PPV, positive predictive value; PR, precision–recall; QC, quality control; ROC, receiver operating characteristic; SBP, systolic blood pressure; SMD, standardized mean difference; TP, true positives; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis; AI, artificial intelligence.

Additional files

File name: Additional file 1.docx. File format: .docx. Title: Supplementary methods and tables.

Additional file 1 contains the following materials: Supplementary Methods S1: Data sources and measurement timing; Supplementary Methods S2: Complete model and preprocessing specifications; Supplementary Methods S3: Implementation and interpretation of N1–N3 and N2V; Supplementary Methods S4: Missingness and selection; Table S1: Predictor missingness by analysis window; Supplementary Methods S5: Additional results; Table S2: Joint calibration parameters for calibrated models; Table S3: Performance in earlier holdout windows; Table S4: Existing sensitivity analyses and unestimated branches; Table S5: Grouped data for calibration plots; Supplementary Methods S6: Original fields and measurement timing; Table S6: Sources and derivation rules for main predictors; Table S7: Exclusions and retained counts at sequential, mutually exclusive eligibility steps; Supplementary Methods S7: Data and code availability; Supplementary Methods S8: Additional definitions and supporting results; Table S8: TRIPOD+AI reporting checklist; Table S9: Algorithm-family comparison in development; Supplementary Methods S9: Analysis chronology; Table S10: Single-window development performance; Table S11: Characteristics by CysC availability; Table S12: Characteristics by outcome observation; Supplementary Methods S10: Exploratory methodological extensions; Table S13: Additional measurements and quality control; Table S14: Extended-model performance; Table S15: Paired increments across the clinical sequence; Table S16: Fixed-capacity assessment comparisons; Table S17: Single-window development crossed with common regularization; Table S18: Complete outer-fold evaluation and calibration records; Table S19: State-specific probability diagnostics.

Declarations

Ethics approval and consent to participate

This study used publicly available, de-identified CHARLS data obtained through the study data-access procedures for a secondary analysis of cystatin C and care needs in older adults. It involved no new participant recruitment, intervention, or collection of identifiable information. The CHARLS protocols were approved by the Biomedical Ethics Review Committee of Peking University (household survey and anthropometric measurements: IRB00001052-11015; blood biomarker collection: IRB00001052-11014) [22,50]. Additional ethics review was considered exempt under Article 32(1) and (2) of China's Measures for the Ethical Review of Life Science and Medical Research Involving Human Participants (2023), which permits exemption for eligible research using legally obtained public data or anonymized information when the specified safeguards concerning harm, sensitive personal information, and commercial interests are met [51]. No separate ethics approval or committee-issued exemption was obtained for this secondary analysis. Data were used in accordance with the CHARLS data-use requirements. The research was conducted in accordance with the ethical principles of the Declaration of Helsinki and the International Ethical Guidelines for Health-related Research Involving Humans of the Council for International Organizations of Medical Sciences [52,53].

Written informed consent was obtained from CHARLS participants [22,50].

Consent for publication

Not applicable.

Availability of data and materials

Frozen model parameters, aggregate data supporting the tables and figures, and the complete analytical code for data construction, model development, model application, and performance evaluation are available at Zenodo (https://doi.org/10.5281/zenodo.23297466). Participant-level CHARLS data are accessed through the CHARLS data-access procedures (https://charls.pku.edu.cn/).

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

Formal analysis, Y.L.; validation, Q.Y., Q.Z. and L.Y.; writing—original draft preparation, Y.L. and R.L.; writing—review and editing, Y.L. and R.L. Y.L. and R.L. contributed equally to this work. All authors have read and agreed to the submitted version of the manuscript.

Acknowledgements

The authors used OpenAI GPT-6 Astra tools to assist with English language editing. The authors reviewed the edited text and take responsibility for the final manuscript.

Study Registration, Protocol and Patient and Public Involvement

The study was not registered. Dated analysis specifications are available in the Zenodo repository, and the analysis chronology is summarized in Supplementary Methods S9. Patients and members of the public were not involved in this secondary analysis.

Clinical trial number

not applicable.

References

1. Chen X, Giles J, Yao Y, Yip W, Meng Q, Berkman L, Chen H, Chen X, Feng J, Feng Z, et al. The path to healthy ageing in China: a Peking University-Lancet Commission. *Lancet.* **2022**;*400*:1967–2006. https://doi.org/10.1016/s0140-6736(22)01546-x.

2. Lv Y, Fan L, Zhou J, Ding E, Shen J, Tang S, He Y, Shi X. Burden of non-communicable diseases due to population ageing in China: challenges to healthcare delivery and long term care services. *BMJ.* **2024**;*387*:e076529. https://doi.org/10.1136/bmj-2023-076529.

3. Gong J, Wang G, Wang Y, Chen X, Chen Y, Meng Q, Yang P, Yao Y, Zhao Y. Nowcasting and forecasting the care needs of the older population in China: analysis of data from the China Health and Retirement Longitudinal Study (CHARLS). *Lancet Public Health.* **2022**;*7*:e1005–e1013. https://doi.org/10.1016/s2468-2667(22)00203-1.

4. Volkert D, Beck AM, Cederholm T, Cruz-Jentoft A, Hooper L, Kiesswetter E, Maggio M, Raynaud-Simon A, Sieber C, Sobotka L, et al. ESPEN practical guideline: Clinical nutrition and hydration in geriatrics. *Clin Nutr.* **2022**;*41*:958–989. https://doi.org/10.1016/j.clnu.2022.01.024.

5. Zhou JH, Lyu YB, Wei Y, Wang JN, Ye LL, Wu B, Liu Y, Qiu YD, Zheng XL, Guo YB, et al. [Prediction of 6-year risk of activities of daily living disability in elderly aged 65 years and older in China]. *Zhonghua Yi Xue Za Zhi.* **2022**;*102*:94–100. (In Chinese) https://doi.org/10.3760/cma.j.cn112137-20210706-01512.

6. Yu Q, Li Z, Yang C, Zhang L, Xing M, Li W. Predicting functional dependency using machine learning among a middle-aged and older Chinese population. *Arch Gerontol Geriatr.* **2023**;*115*:105124. https://doi.org/10.1016/j.archger.2023.105124.

7. Han Y, Wang S. Disability risk prediction model based on machine learning among Chinese healthy older adults: results from the China Health and Retirement Longitudinal Study. *Front Public Health.* **2023**;*11*:1271595. https://doi.org/10.3389/fpubh.2023.1271595.

8. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, Crews DC, Doria A, Estrella MM, Froissart M, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. *N Engl J Med.* **2021**;*385*:1737–1749. https://doi.org/10.1056/nejmoa2102953.

9. Pottel H, Björk J, Rule AD, Ebert N, Eriksen BO, Dubourg L, Vidal-Petiot E, Grubb A, Hansson M, Lamb EJ, et al. Cystatin C-Based Equation to Estimate GFR without the Inclusion of Race and Sex. *N Engl J Med.* **2023**;*388*:333–343. https://doi.org/10.1056/nejmoa2203769.

10. Chen DC, Potok OA, Rifkin D, Estrella MM. Advantages, Limitations, and Clinical Considerations in Using Cystatin C to Estimate GFR. *Kidney360.* **2022**;*3*:1807–1814. https://doi.org/10.34067/kid.0003202022.

11. Li C, Ma Y, Yang C, Hua R, Xie W, Zhang L. Association of Cystatin C Kidney Function Measures With Long-term Deficit-Accumulation Frailty Trajectories and Physical Function Decline. *JAMA Netw Open.* **2022**;*5*:e2234208. https://doi.org/10.1001/jamanetworkopen.2022.34208.

12. Bai A, Xu J, Xu W, Cao J, Zhao B. Creatinine and cystatin C-based indices for predicting sarcopenia, frailty and disability in older community-dwelling adults. *J Nutr Health Aging.* **2025**;*29*:100635. https://doi.org/10.1016/j.jnha.2025.100635.

13. Wu S, Wang F, Feng X, Xin S, Li X, Zhao J, Li S, Meng Z, Chen H, Yang Y, et al. Association Between Serum Creatinine to Cystatin C Ratio and Activities of Daily Living Disability Trajectories: Evidence From the China Health and Retirement Longitudinal Study. *Geriatr Gerontol Int.* **2026**;*26*:e70683. https://doi.org/10.1111/ggi.70683.

14. Deng L, Zheng X, Chen Y, Liu C, Shi J, Bu Z, Liu X, Zhao H, Li S, Yin B, et al. The Predictive Power of the Cystatin C-Creatinine Score in Assessing Frailty. *J Cachexia Sarcopenia Muscle.* **2025**;*16*:e70040. https://doi.org/10.1002/jcsm.70040.

15. Matsuzawa R, Nagai K, Mori T, Onishi M, Tsuji S, Hashimoto K, Tamaki K, Wada Y, Kusunoki H, Nagasawa Y, et al. A serum creatinine-cystatin C-derived index as a predictive marker for sarcopenia incidence and progression in community-dwelling older adults: a prospective cohort study. *Age Ageing.* **2025**;*54*:afaf185. https://doi.org/10.1093/ageing/afaf185.

16. Steyerberg EW, Pencina MJ, Lingsma HF, Kattan MW, Vickers AJ, Van Calster B. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Investig.* **2012**;*42*:216–228. https://doi.org/10.1111/j.1365-2362.2011.02562.x.

17. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* **2010**;*21*:128–138. https://doi.org/10.1097/ede.0b013e3181c30fb2.

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

19. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak.* **2006**;*26*:565–574. https://doi.org/10.1177/0272989x06295361.

20. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). *Int J Epidemiol.* **2014**;*43*:61–68. https://doi.org/10.1093/ije/dys203.

21. Zhao Y, Crimmins E, Hu P, Hu Y, Ge T, Kim JK, Strauss J, Yang G, Yin X, Wang Y. *China Health and Retirement Longitudinal Study 2011–2012 National Baseline Blood Data Users' Guide*. Beijing, China: China Center for Economic Research, Peking University; 2014. Available from: https://charls.charlsdata.com/Public/ashelf/public/uploads/document/2011-charls-wave1/application/blood_user_guide_en_20140429.pdf Accessed 11 Oct 2026.

22. Chen X, Crimmins E, Hu PP, Kim JK, Meng Q, Strauss J, Wang Y, Zeng J, Zhang Y, Zhao Y. Venous Blood-Based Biomarkers in the China Health and Retirement Longitudinal Study: Rationale, Design, and Results From the 2015 Wave. *Am J Epidemiol.* **2019**;*188*:1871–1877. https://doi.org/10.1093/aje/kwz170.

23. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* **2024**;*385*:e078378. https://doi.org/10.1136/bmj-2023-078378.

24. Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB, Moons KGM. Transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster): explanation and elaboration. *BMJ.* **2023**;*380*:e071058. https://doi.org/10.1136/bmj-2022-071058.

25. Ding H, Wang K, Li Y, Zhao X. Trends in disability in activities of daily living and instrumental activities of daily living among Chinese older adults from 2011 to 2018. *Aging Clin Exp Res.* **2024**;*36*:27. https://doi.org/10.1007/s40520-023-02690-7.

26. Gehringer CK, Martin GP, Van Calster B, Hyrich KL, Verstappen SMM, Sergeant JC. How to develop, validate, and update clinical prediction models using multinomial logistic regression. *J Clin Epidemiol.* **2024**;*174*:111481. https://doi.org/10.1016/j.jclinepi.2024.111481.

27. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res.* **2019**;*3*:18. https://doi.org/10.1186/s41512-019-0064-7.

28. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, Collins GS. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *J Clin Epidemiol.* **2021**;*132*:88–96. https://doi.org/10.1016/j.jclinepi.2020.12.005.

29. Potok OA, Ix JH, Shlipak MG, Bansal N, Katz R, Kritchevsky SB, Rifkin DE. Cystatin C- and Creatinine-Based Glomerular Filtration Rate Estimation Differences and Muscle Quantity and Functional Status in Older Adults: The Health, Aging, and Body Composition Study. *Kidney Med.* **2022**;*4*:100416. https://doi.org/10.1016/j.xkme.2022.100416.

30. Tabara Y, Okada Y, Ochi M, Ohyagi Y, Igase M. Association of Creatinine-to-Cystatin C Ratio with Myosteatosis and Physical Performance in Older Adults: The Japan Shimanami Health Promoting Program. *J Am Med Dir Assoc.* **2021**;*22*:2366–2372.e3. https://doi.org/10.1016/j.jamda.2021.03.021.

31. Katz-Agranov N, Rieu-Werden ML, Thacker A, Lykken JM, Sise ME, Shah SJ. Large Discordance between Creatinine-Based and Cystatin C-Based eGFRs is Associated with Falls, Hospitalizations, and Death in Older Adults. *Clin J Am Soc Nephrol.* **2024**;*19*:1275–1283. https://doi.org/10.2215/cjn.0000000000000523.

32. Xu L, Xu W, Qin L. Association of cystatin C kidney function measures with motoric cognitive risk syndrome: evidence from two cohort studies. *J Nutr Health Aging.* **2025**;*29*:100484. https://doi.org/10.1016/j.jnha.2025.100484.

33. Osawa Y, Sasaki T, Candia J, Abe Y, Wong G, Bandinelli S, Ferrucci L, Arai Y. Plasma proteins associated with disability and mortality risks in Japanese community-dwelling octogenarians. *GeroScience.* **2026**. https://doi.org/10.1007/s11357-026-02377-7.

34. Writing Group for the CKD Prognosis Consortium. Estimated Glomerular Filtration Rate, Albuminuria, and Adverse Outcomes: An Individual-Participant Data Meta-Analysis. *JAMA.* **2023**;*330*:1266–1277. https://doi.org/10.1001/jama.2023.17002.

35. Estrella MM, Ballew SH, Sang Y, Grams ME, Coresh J, Surapaneni A, Alencar de Pinho N, Ärnlöv J, Brenner H, Carrero JJ, et al. Discordance in Creatinine- and Cystatin C-Based eGFR and Clinical Outcomes: A Meta-Analysis. *JAMA.* **2025**;*334*:1915–1926. https://doi.org/10.1001/jama.2025.17578.

36. McCoy IE, Yang J, Go AS, Rincon-Choles H, Taliercio J, Rosas SE, Unruh M, Shah V, Cohen DL, He J, et al. Complex Etiologies of the Discordance Between Cystatin C- and Creatinine-Based Estimated GFR and Its Adverse Associations: Findings From the CRIC Study. *Am J Kidney Dis.* **2025**;*86*:192–201. https://doi.org/10.1053/j.ajkd.2025.03.018.

37. Shi Z, Zhang Z, Shi K, Yu B, Jiang Z, Yang L, Lin J, Fang Y. Association between multimorbidity trajectories and incident disability among mid to older age adults: China Health and Retirement Longitudinal Study. *BMC Geriatr.* **2022**;*22*:741. https://doi.org/10.1186/s12877-022-03421-9.

38. Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, Bonnett LJ, Collins GS. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. *BMJ.* **2024**;*384*:e074820. https://doi.org/10.1136/bmj-2023-074820.

39. Crocker TF, Ensor J, Lam N, Jordão M, Bajpai R, Bond M, Forster A, Riley RD, Andre D, Brundle C, 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.

40. Lin Z, Qian Y, Gill TM, Hou X, Allore H, Chen S, Chen X. Absence of care among community-dwelling older adults with dementia and functional limitations. *Nat Aging.* **2025**;*5*:548–557. https://doi.org/10.1038/s43587-025-00836-y.

41. Cederholm T, Jensen GL, Correia MITD, Gonzalez MC, Fukushima R, Pisprasert V, Blaauw R, Braz DC, Carrasco F, Cruz Jentoft AJ, et al. The GLIM consensus approach to diagnosis of malnutrition: A 5-year update. *Clin Nutr.* **2025**;*49*:11–20. https://doi.org/10.1016/j.clnu.2025.03.018.

42. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, Beam AL, Van Calster B, Celi LA, Denaxas S, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* **2025**;*388*:e082505. https://doi.org/10.1136/bmj-2024-082505.

43. Martin GP, Riley RD, Collins GS, Sperrin M. Developing clinical prediction models when adhering to minimum sample size recommendations: The importance of quantifying bootstrap variability in tuning parameters and predictive performance. *Stat Methods Med Res.* **2021**;*30*:2545–2561. https://doi.org/10.1177/09622802211046388.

44. Sperrin M, Riley RD, Collins GS, Martin GP. Targeted validation: validating clinical prediction models in their intended population and setting. *Diagn Progn Res.* **2022**;*6*:24. https://doi.org/10.1186/s41512-022-00136-8.

45. Sisk R, Sperrin M, Peek N, van Smeden M, Martin GP. Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: A simulation study. *Stat Methods Med Res.* **2023**;*32*:1461–1477. https://doi.org/10.1177/09622802231165001.

46. Zhou J, Wei MY, Zhang J, Liu H, Wu C. Association of multimorbidity patterns with incident disability and recovery of independence among middle-aged and older adults. *Age Ageing.* **2022**;*51*:afac177. https://doi.org/10.1093/ageing/afac177.

47. Davies LE, Mercer SW, Brittain K, Jagger C, Robinson L, Kingston A. The association between multimorbidity and mobility disability-free life expectancy in adults aged 85 years and over: A modelling study in the Newcastle 85+ cohort. *PLoS Med.* **2022**;*19*:e1004130. https://doi.org/10.1371/journal.pmed.1004130.

48. Riley RD, Snell KIE, Archer L, Ensor J, Debray TPA, van Calster B, van Smeden M, Collins GS. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. *BMJ.* **2024**;*384*:e074821. https://doi.org/10.1136/bmj-2023-074821.

49. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, Harrell FE, Martin GP, Moons KGM, van Smeden M, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ.* **2024**;*384*:e074819. https://doi.org/10.1136/bmj-2023-074819.

50. Zhao Y, Strauss J, Chen X, Wang Y, Gong J, Meng Q, Wang G, Wang H. *China Health and Retirement Longitudinal Study Wave 4 User's Guide*. Beijing, China: National School of Development, Peking University; 2020. Available from: https://charls.pku.edu.cn/en/data/User2018.pdf Accessed 11 Oct 2026.

51. National Health Commission, Ministry of Education, Ministry of Science and Technology, National Administration of Traditional Chinese Medicine. Measures for the Ethical Review of Life Science and Medical Research Involving Human Participants Beijing, China: National Health Commission; 2023. Available from: https://www.nhc.gov.cn/wjw/c100375/202302/902b4a1dc3af4aba862a6387e6e376dc.shtml Accessed 11 Oct 2026. (In Chinese)

52. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA.* **2025**;*333*:71–74. https://doi.org/10.1001/jama.2024.21972.

53. Council for International Organizations of Medical Sciences (CIOMS). *International Ethical Guidelines for Health-related Research involving Humans*, 4th ed.. Geneva, Switzerland: Council for International Organizations of Medical Sciences; 2016. https://doi.org/10.56759/rgxl7405.

Supplementary Files

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

- Additionalfile1.docx