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.

Nationally implemented routine-data framework stratifies care dependency risk by multidimensional frailty-related vulnerability in older adults

Tomoki Tanaka
The University of Tokyo
Yasuyo Yoshizawa
Juntendo University
Kazuyo Tsushita
Kagawa Nutrition University
Katsuya Iijima
iijima@iog.u-tokyo.ac.jp

The University of Tokyo

Article

Keywords: frailty, multidomain vulnerability, administrative data, care dependency, risk stratification, healthy aging

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11176983/v1

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

Additional Declarations: No competing interests reported.

Abstract

Frailty prevention at population scale requires integrating functional and medical vulnerability using routinely collected data. We evaluated a nationally implemented Japanese routine-data framework linking physical, oral, nutritional, and medical-support needs in 532,155 adults aged ≥ 75 years. Incident long-term care insurance certification within 365 days ranged from 2.9% in those with no identified needs to 13.9% in those with at least two functional domains plus medical needs. The adjusted odds ratio for the highest-need group was 2.56 (95% confidence interval, 2.34–2.80). A combined model showed stable later-period discrimination (area under the receiver operating characteristic curve, 0.801) but underestimated absolute risk, supporting local recalibration. Repeated assessments also updated prognosis: worsening functional or medical status was associated with higher risk, whereas several transitions toward lower vulnerability were associated with lower risk. These findings support multidomain routine-data stratification as a scalable complement to direct frailty assessment for prioritizing preventive follow-up in aging populations.

INTRODUCTION

Frailty and care dependency emerge from interacting vulnerabilities across physical function, nutrition, oral health, chronic disease, and treatment burden rather than from a single disease process (1, 2). Population aging therefore creates a practical challenge for health systems: how to identify older adults who may benefit from preventive assessment before disability becomes established. Routine administrative and health check-up data offer a scalable complement to direct frailty assessment because they can be applied across large populations without additional research-specific examinations (3–6).

Japan provides a distinctive setting in which to test this approach. Its universal long-term care insurance (LTCI) system is increasingly linked with preventive health services for older adults (7–11). Within this system, a nationally implemented Kokuho Database (KDB)-based support framework combines routinely collected health check-up, questionnaire, medical claims, medication, and LTCI information to identify functional vulnerability and medical-support needs. The functional component captures reduced physical function, reduced oral function, and undernutrition risk, while medical-support criteria capture chronic-disease and medication-related needs (7, 12–14). These criteria were designed for population case finding and service prioritization, but their prognostic performance and behavior over repeated assessments have not been comprehensively evaluated.

Previous KDB studies have shown that questionnaire-derived frailty is associated with disability and healthcare costs (3, 4). For a multidomain framework intended to inform preventive follow-up, however, evidence is also needed on absolute risk separation, incremental predictive information beyond basic characteristics, calibration over time, operating characteristics at practical thresholds, and whether repeated measurements update prognosis (5, 6, 16). A 365-day horizon is operationally relevant because it captures near-term care dependency within an annual reassessment cycle, when risk information may be most useful for prioritizing preventive follow-up, while limiting the influence of longer-term changes in health status and service use.

We therefore evaluated whether a routine-data framework integrating functional and medical vulnerability stratifies 365-day risk of incident LTCI certification in older adults. We additionally assessed later-period discrimination and calibration, study-defined operating thresholds, transitions between repeated assessments, robustness to an alternative oral-function definition, and contemporaneous medical and long-term care expenditure. The analytic framework is summarized in Fig. 1.

RESULTS

Primary cohort

The primary cohort included 532,155 participants (mean age, 79.91 years; 54.5% women). Functional vulnerability was common: 15.5% met the reduced-physical-function criterion, 11.8% the reduced-oral-function criterion, and 2.6% the undernutrition criterion. Medical-support needs were present in 9.9%. The six-group distribution was 68.8%, 5.2%, 18.3%, 3.9%, 2.9%, and 0.9%, respectively (Table 1).

Table 1
Baseline characteristics of the primary 365-day cohort (N = 532,155)

Characteristic	Value
Basic attributes	
Age, years	79.91 (3.97)
Female sex	289,902 (54.5%)
Body mass index, kg/m²	22.89 (3.22)
Systolic blood pressure, mmHg	135.71 (16.90)
Diastolic blood pressure, mmHg	73.91 (10.56)
HbA1c, %	5.84 (0.98)
eGFR, mL/min/1.73 m²	62.83 (18.25)
Charlson Comorbidity Index	2.11 (1.67)
Current smoking	29,561 (5.6%)
No weekly walking	180,605 (34.1%)
< 3 meals per day	23,894 (4.5%)
Hypertension treatment	329,012 (61.8%)
Diabetes treatment	75,777 (14.2%)
Dyslipidemia treatment	248,657 (46.7%)
Targeting-tool exposures	
Reduced physical function	82,655 (15.5%)
Reduced oral function	62,651 (11.8%)
Undernutrition risk	14,102 (2.6%)
Functional-domain count: 0	393,801 (74.0%)
Functional-domain count: 1	118,260 (22.2%)
Functional-domain count: ≥2	20,094 (3.8%)
Poorly controlled chronic conditions	15,689 (2.9%)
Diabetic nephropathy / renal risk	413 (0.1%)
≥ 15 prescribed medications	12,462 (2.3%)
Hypnotic medication guidance	26,804 (5.0%)
Any medical-support need	52,825 (9.9%)

Note. Values are mean (SD) or n (%) unless a denominator is shown. The program diabetes-frailty criterion was analyzed separately and was not included in the medical-needs axis. eGFR indicates estimated glomerular filtration rate; FY, fiscal year; HbA1c, glycated hemoglobin.

Characteristic	Value
Program diabetes-frailty criterion	103,181 (19.4%)
Index FY2020	359,410 (67.5%)
Index FY2021	129,132 (24.3%)
Index FY2022 (through Aug 31)	43,613 (8.2%)

Note. Values are mean (SD) or n (%) unless a denominator is shown. The program diabetes-frailty criterion was analyzed separately and was not included in the medical-needs axis. eGFR indicates estimated glomerular filtration rate; FY, fiscal year; HbA1c, glycated hemoglobin.

Six-group prognostic stratification

Incident LTCI certification within 365 days occurred in 21,821 participants (4.1%). Event rates across the six categories were 2.9% in those with no functional or medical needs, 5.3% with medical needs only, 5.7% with 1 functional domain only, 9.8% with 1 functional domain plus medical needs, 9.2% with ≥ 2 functional domains only, and 13.9% with ≥ 2 functional domains plus medical needs. In complete-case models, adjusted ORs relative to no functional or medical needs were 1.44, 1.51, 2.03, 2.05, and 2.56, respectively (all P<.001; Table 2). Associations for individual targeting criteria were heterogeneous but generally in the expected direction; stronger associations were observed for reduced physical function, diabetic nephropathy, hypnotic medication guidance, and accumulation of functional domains (Supplementary eTable 3).

Table 2
Incident 365-day LTCI certification according to the 3 × 2 functional-medical classification

Group	n (%)	Events, n (%)	Adjusted OR (95% CI)	*P value*
No functional or medical needs	366,337 (68.8)	10,698 (2.9)	1.000 (reference)	—
Medical needs only	27,464 (5.2)	1,458 (5.3)	1.435 (1.354–1.521)	*< .001*
1 functional domain only	97,632 (18.3)	5,573 (5.7)	1.514 (1.463–1.568)	*< .001*
1 functional domain + medical needs	20,628 (3.9)	2,017 (9.8)	2.031 (1.927–2.141)	*< .001*
≥ 2 functional domains only	15,361 (2.9)	1,418 (9.2)	2.050 (1.928–2.180)	*< .001*
≥ 2 functional domains + medical needs	4,733 (0.9)	657 (13.9)	2.555 (2.335–2.797)	*< .001*

Note. CI indicates confidence interval; LTCI, long-term care insurance; OR, odds ratio. Crude group sizes and event counts use the full primary cohort (N = 532,155). Adjusted ORs use complete cases (n = 528,042; 21,619 events) and are adjusted for age, sex, body mass index, Charlson Comorbidity Index, current smoking, no weekly walking, < 3 meals/day, and fiscal year. The six groups are nominal categories, not an ordinal score.

Prediction and operational performance

The development sample included 484,818 participants (19,710 events), and the later-period validation sample included 43,224 participants (1,909 events). Development AUCs were 0.753 (95% CI, 0.750–0.757) for the basic

model, 0.615 (0.610–0.619) for the tool-only model, and 0.764 (0.761–0.767) for the combined model. Validation AUCs were 0.794 (0.784–0.804), 0.614 (0.600–0.628), and 0.801 (0.791–0.811), respectively (Table 3). Validation Brier scores were 0.0397, 0.0417, and 0.0397. For the combined model, mean predicted risk was 3.42% versus an observed risk of 4.42%; the calibration slope was 1.094 (95% CI, 1.047–1.141) and calibration-in-the-large was 0.285 (95% CI, 0.237–0.332). Full model coefficients and detailed performance estimates are provided in Supplementary eTable 4. Decile-based calibration showed slight overprediction at the lowest risks and increasing underprediction in the upper deciles (Supplementary eFigure 2).

Table 3
Development and later-period validation of 365-day LTCI prediction and operational thresholds

Panel A. Model performance	**Sample**	**AUC (95% CI)**	**Brier score**		**Mean predicted risk**
Basic model	Development	0.753 (0.750–0.757)	0.0374		4.07%
Tool-only model	Development	0.615 (0.610–0.619)	0.0386		4.07%
Combined model	Development	0.764 (0.761–0.767)	0.0373		4.07%
Basic model	Validation	0.794 (0.784–0.804)	0.0397		3.40%
Tool-only model	Validation	0.614 (0.600–0.628)	0.0417		4.03%
Combined model	Validation	0.801 (0.791–0.811)	0.0397		3.42%
Panel B. Operational threshold in validation	Flagged, n (%)	Sensitivity	Specificity	PPV	NPV
Any need (groups 2–6)	13,177 (30.5%)	50.1%	70.4%	7.3%	96.8%
Higher-need categories (groups 4–6)	3,146 (7.3%)	18.8%	93.3%	11.4%	96.1%
Note. Operational thresholds were defined for this study and are not official program cutoffs. AUC indicates area under the receiver operating characteristic curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.					
Development: n = 484,818; 19,710 events (4.07%). Validation: n = 43,224; 1,909 events (4.42%). In validation, the combined model calibration slope was 1.094 (95% CI, 1.047–1.141) and calibration-in-the-large was 0.285 (95% CI, 0.237–0.332).					
Basic model: age, sex, BMI, CCI, current smoking, no weekly walking, and < 3 meals/day. Tool-only model: six-group classification. Combined model: basic variables plus the six-group classification.					

Operational thresholds

In validation, the any-need threshold flagged 30.5% of participants and yielded sensitivity 50.1%, specificity 70.4%, positive predictive value (PPV) 7.3%, and negative predictive value (NPV) 96.8%. The higher-need threshold flagged 7.3% and yielded sensitivity 18.8%, specificity 93.3%, PPV 11.4%, and NPV 96.1% (Table 3).

Transitions

The transition cohort included 309,123 complete cases and 14,582 incident LTCI events. Among participants with no functional domains at the previous assessment, transition to 1 domain (adjusted OR, 1.64; 95% CI, 1.55–1.72) or ≥ 2 domains (2.33; 2.02–2.68) was associated with higher subsequent risk than remaining at 0. Among participants previously at 1 domain, transition to 0 was associated with lower risk (0.79; 0.74–0.84), whereas transition to ≥ 2 was associated with higher risk (1.25; 1.12–1.38). Among participants previously at ≥ 2 domains, transition to 1 was associated with lower risk (0.77; 0.66–0.89), whereas transition directly to 0 was not statistically significant (0.87; 0.71–1.06). New onset of medical needs was associated with higher risk than remaining without medical needs (1.51; 1.42–1.61), and persistent medical needs were associated with higher risk than resolution (1.19; 1.09–1.31; Table 4).

Table 4
Associations of transitions in functional and medical needs with subsequent 365-day LTCI certification

Previous status → current status	n	Events, n (%)	Adjusted OR (95% CI)	*P value*
Functional domains: previous 0				
0 → 0	201,154	6,560 (3.3)	1.000 (reference)	—
0 → 1	31,176	2,152 (6.9)	1.635 (1.552–1.723)	*< .001*
0 → ≥2	2,262	254 (11.2)	2.327 (2.022–2.679)	*< .001*
Functional domains: previous 1				
1 → 0	27,407	1,546 (5.6)	0.786 (0.735–0.841)	*< .001*
1 → 1	32,631	2,528 (7.7)	1.000 (reference)	—
1 → ≥2	4,724	503 (10.6)	1.245 (1.120–1.383)	*< .001*
Functional domains: previous ≥ 2				
≥ 2 → 0	1,576	154 (9.8)	0.867 (0.712–1.055)	*.155*
≥ 2 → 1	4,275	392 (9.2)	0.766 (0.663–0.885)	*< .001*
≥ 2 → ≥2	3,918	493 (12.6)	1.000 (reference)	—
Medical needs: previous absent				
No need → no need	263,484	10,967 (4.2)	1.000 (reference)	—
No need → onset	17,268	1,401 (8.1)	1.511 (1.423–1.606)	*< .001*
Medical needs: previous present				
Need → resolution	15,773	1,035 (6.6)	1.000 (reference)	—
Need → persistent	12,598	1,179 (9.4)	1.194 (1.090–1.307)	*< .001*

Note. Complete-case transition cohort n = 309,123; 14,582 events. Each adjusted comparison is conditional on previous status and adjusts for age, sex, BMI, CCI, current smoking, no weekly walking, < 3 meals/day, interval between assessments, and current fiscal year. CI indicates confidence interval; LTCI, long-term care insurance; OR, odds ratio.

Sensitivity and expenditure analyses

Results were stable when the administrative cutoff was extended to September 30, 2022 (n = 551,760; 22,557 events); adjusted ORs for groups 2–6 were 1.43, 1.52, 2.04, 2.06, and 2.55. Classifying all initial certification dates within 365 days as events regardless of missing initial care level also produced similar gradients. In an additional sensitivity analysis, the dental-visit component was excluded and reduced oral function was defined using QMCOO item 4 or 5 alone. Among 526,567 complete cases with 21,539 events, adjusted ORs for groups

2–6 were 1.45, 1.18, 1.73, 1.80, and 2.18, respectively. In later-period validation (n = 43,166; 1,907 events), AUCs were 0.794, 0.610, and 0.801 for the basic, tool-only, and combined models, respectively; corresponding Brier scores were 0.0398, 0.0418, and 0.0397 (Supplementary eTable 5). Mean contemporaneous annual medical and LTC expenditure was JPY 401,999 (median, JPY 250,180; interquartile range, JPY 140,100–433,430), and 1.5% had zero expenditure. Relative to the no-needs group, adjusted cost ratios were 1.10, 1.08, 1.30, 1.14, and 1.31 across groups 2–6 (Supplementary eTable 6).

DISCUSSION

In this large linked-data cohort, multidomain vulnerability derived from routine data stratified near-term care dependency risk at population scale. The main advance is not a stand-alone high-performance prediction model, but the demonstration that an existing care-system framework integrating physical, oral, nutritional, and medical signals identifies clinically distinct risk strata, retains prognostic information in a later period, and can be updated with repeated assessment. Crude 365-day event rates differed nearly fivefold between the lowest- and highest-risk groups (2.9% versus 13.9%). At the same time, the modest incremental gain in discrimination after adding the six-group classification to basic covariates (AUC 0.794 to 0.801) argues for using multidomain vulnerability as an operational layer within a broader assessment strategy rather than as a stand-alone prediction model.

The functional component is consistent with established pathways linking mobility limitation, oral dysfunction, and undernutrition to disability and frailty (1, 17–22). Accumulation across functional domains was more informative than any single domain, supporting a multidomain interpretation. Medical-support criteria contributed additional stratification in several cells of the 3×2 matrix, particularly when functional vulnerability was also present. Medication-related and renal-risk indicators plausibly capture multimorbidity, treatment complexity, and reduced physiologic reserve (23–26). These mechanisms are consistent with prognostic association but do not establish that changing a criterion would itself alter subsequent LTCI risk.

The calibration results are operationally important. In later-period validation, the combined model underpredicted average risk (3.42% predicted vs 4.42% observed), and underprediction became more pronounced in higher-risk deciles. This pattern argues against direct use of development-sample probabilities as fixed national risk estimates. Local recalibration or periodic updating would be appropriate if absolute predicted risks are used for decision thresholds. Discrimination was nevertheless maintained in the later-period sample, which is consistent with using the model primarily for relative prioritization rather than uncalibrated absolute-risk decisions.

The threshold analyses illustrate the trade-off between reach and concentration. Flagging any need identified about half of subsequent LTCI events but had a PPV of 7.3%; restricting attention to higher-need categories increased specificity and PPV but identified fewer than one in five events. These are not diagnostic thresholds, and we did not evaluate decision-curve net benefit or utilities associated with specific outreach actions. Accordingly, these metrics should be interpreted descriptively as operating characteristics rather than as evidence that a particular threshold is optimal. They may nonetheless inform how municipalities consider balancing broader reach against concentration of higher-risk individuals, while recognizing that many events arise outside flagged categories. Population screening linked to differentiated geriatric follow-up has been implemented in integrated-care settings (27), and claims-based frailty indices likewise demonstrate the feasibility of scalable risk stratification using routinely collected data (28). In Japan, senior volunteer-led Frailty

Check results have predicted subsequent disability and mortality, while a 10-year population-based matched study found programme participation associated with lower LTCI certification and LTC expenditures (29, 30). The KDB tool evaluated here may complement such face-to-face programmes by informing whom municipalities consider for further assessment; however, the present study did not test whether KDB-guided outreach or referral improves outcomes.

The transition analysis further supports repeated assessment as a prognostic process. Transitions to more functional domains and new medical needs were associated with higher subsequent risk within strata of prior status, whereas several transitions to fewer functional domains were associated with lower risk. This pattern is consistent with longitudinal evidence that changes in frailty status are associated with subsequent disability (18). Because the analysis required a second eligible assessment and absence of known LTCI certification at the landmark, it selected participants who remained under observation and returned for reassessment; associations involving apparent improvement or resolution may therefore be affected by return-attendance selection. Because these are observational transitions, they should not be interpreted as effects of treatment or recovery. Rather, the findings indicate that repeated routine assessments can update prognostic information as functional and medical status changes over time among participants who return for reassessment.

The expenditure analysis provides complementary service-use context but has different temporality from the primary outcome. Annual expenditure was contemporaneous with the extraction fiscal year, not a prospective 365-day cost endpoint. The classification also differentiated contemporaneous resource use after adjustment, but these estimates should not be interpreted as future cost prediction or economic effects. The right-skewed distribution and nontrivial zero mass supported the Tweedie modeling approach used here.

This study has several limitations. First, the cohort consisted of adults who attended health check-ups, which may limit generalizability to less-engaged or frailer populations. Data were drawn from municipalities in two prefectures, and performance may differ in other geographic or administrative settings. The study period also overlapped the COVID-19 pandemic, which may have affected health check-up attendance, care-seeking, LTCI application or certification patterns, and healthcare utilization; adjustment for fiscal year may not fully capture these period effects. Second, individual-level dates and reasons for loss of insurance eligibility, including death, were unavailable. Participants who died before 365 days without LTCI certification may therefore have been classified as non-events, and formal competing-risk analyses were not feasible. Third, because incident LTCI was counted from day 1 after the index assessment, some early certifications may reflect functional deterioration already present at baseline rather than longer-horizon prognostic separation. In addition, 2,292 participants had an initial certification date within 365 days but lacked an initial care level; sensitivity analyses that treated these dates as events produced similar results. Fourth, the later-period validation sample represented only the portion of FY2022 with adequate administrative outcome availability and should be interpreted as later-period temporal validation rather than full-fiscal-year or geographic external validation. Fifth, the targeting criteria are operational program rules rather than adjudicated clinical diagnoses. In particular, the 12-month dental-visit component of the oral-function criterion used a program-defined reference period that was not individually anchored to the index assessment date, creating potential temporal misalignment. However, both the six-group risk gradient and prediction performance were preserved in a sensitivity analysis that excluded the dental-visit component and defined oral function using QMCOO items 4 or 5 alone. Residual confounding is also possible. The transition analysis additionally required return for a second eligible assessment and absence of known LTCI certification at the landmark, so associations involving apparent improvement or resolution may be affected by selection

among return attendees. Finally, expenditure analyses were contemporaneous and excluded dental costs because of substantial missingness.

Taken together, these findings show that routinely collected data can operationalize multidomain vulnerability for population-scale stratification of one-year care dependency risk. The framework complements rather than replaces direct frailty assessment: it can help identify groups for differentiated follow-up, while absolute predicted risks require local recalibration and threshold choice should reflect local resources and intervention goals. Repeated assessment adds prognostic information as functional and medical status change over time. This approach may be relevant to aging health systems seeking scalable links between frailty detection, preventive outreach, and long-term care planning.

METHODS

Study design and data sources

This retrospective cohort study used de-identified linked administrative and health check-up data from municipalities in Kanagawa and Aichi prefectures, Japan. Data included annual health check-up measurements, the 15-item Questionnaire for Medical Checkup of Old-Old (QMCOO), KDB-derived medical and pharmacy claims indicators, and LTCI records (8, 9, 12, 15). The source analytic file contained 1,256,250 annual assessment records from 630,300 unique individuals aged ≥ 75 years during fiscal years (FYs) 2020–2022.

Primary cohort and index date

For each individual, the earliest eligible health check-up was selected as the index assessment. To provide an administrative 365-day ascertainment window, the primary cohort was restricted to index dates on or before August 31, 2022. Participants with LTCI certification on the index date were excluded to preserve temporal ordering. The resulting primary cohort included 532,155 participants. A sensitivity analysis extended the administrative cutoff to September 30, 2022. Cohort construction is shown in Supplementary eFigure 1.

Targeting-tool exposures

Functional vulnerability comprised three binary domains: reduced physical function, reduced oral function, and undernutrition risk, operationalized according to nationally specified support-tool rules (7, 12–14). A functional-domain count was categorized as 0, 1, or ≥ 2. Medical-support needs were defined as the presence of at least one of four program criteria: poorly controlled chronic conditions, diabetic nephropathy/renal risk, ≥ 15 prescribed medications, or hypnotic medication guidance. The program diabetes-frailty criterion was evaluated separately and was not included in the medical-needs axis. We cross-classified functional-domain count with medical-support needs (absent/present) to create six mutually exclusive groups: no functional or medical needs; medical needs only; 1 functional domain only; 1 functional domain plus medical needs; ≥2 functional domains only; and ≥ 2 functional domains plus medical needs. The six categories were treated as nominal rather than as an ordinal score. The reduced-oral-function criterion combined QMCOO item 4 or 5 with a claims-based dental-visit criterion defined over a program-specific 12-month reference period. Because this reference period was not individually anchored to each participant's index assessment date, we conducted an additional sensitivity analysis in which reduced oral function was defined using QMCOO item 4 or 5 alone, excluding the dental-visit component. Detailed operational definitions are provided in Supplementary eTables 1–2.

Outcomes

The primary outcome was incident LTCI certification from 1 through 365 days after the index assessment, including Support Level 1 through Care Level 5. Participants with an initial certification date but missing initial care level were classified as non-events in the primary definition and included as events in a sensitivity analysis. Care Level 1 or higher and Care Level 2 or higher were examined as severity-oriented secondary outcomes. Individual-level dates and reasons for loss of eligibility, including death, were unavailable in the analytic dataset. Therefore, death could not be treated as a competing event and formal competing-risk analyses were not feasible.

Covariates

Prespecified covariates were age, sex, body mass index (BMI), Charlson Comorbidity Index (CCI), current smoking, no weekly walking, fewer than three meals per day, and fiscal year. Analyses used complete cases for these covariates; 528,042 of 532,155 participants (99.2%) were complete. No imputation was performed.

Statistical analysis

All analyses were performed in SPSS version 29.0 (IBM Corp., Armonk, NY, USA). Two-sided $P<.05$ was considered statistically significant. Prediction analyses and reporting were guided by TRIPOD + AI principles (16).

Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) for the six-group classification were estimated using multivariable logistic regression. For prediction analyses, complete cases from FY2020–FY2021 formed the development sample (n = 484,818), and the later FY2022 period available under the primary cutoff formed the validation sample (n = 43,224). Three models were estimated in development and then frozen: a basic model (age, sex, BMI, CCI, smoking, no weekly walking, and < 3 meals/day), a tool-only model (six-group classification), and a combined model. Discrimination was summarized using the area under the receiver operating characteristic curve (AUC), overall accuracy by the Brier score, and calibration using the calibration slope, calibration-in-the-large (intercept with slope fixed at 1), and deciles of predicted risk. Two study-defined operational thresholds were examined in validation: any need (groups 2–6) and higher-need categories (groups 4–6). Decision-curve or net-benefit analysis was not performed; threshold results were intended to describe operating characteristics rather than identify an optimal intervention rule.

In the oral-function sensitivity analysis, the functional-domain count and six-group classification were reconstructed using the questionnaire-only definition. Association and prediction analyses were then repeated using the same covariate adjustment, development/validation split, and model specifications as in the primary analyses.

A secondary transition analysis used one eligible adjacent assessment pair per person, required 180–550 days between assessments, and excluded participants with known LTCI certification before or at the landmark assessment. The cohort included 310,999 participants (309,123 complete cases). Functional transitions were analyzed within strata of the previous functional-domain count (0, 1, or ≥ 2), and medical transitions were analyzed within strata of previous medical-need status. Models adjusted for age, sex, BMI, CCI, smoking, no weekly walking, < 3 meals/day, interval between assessments, and current fiscal year.

As a secondary resource-use analysis, annual expenditure in the extraction fiscal year was calculated as inpatient medical, outpatient medical, pharmacy, and LTC benefit expenditure in Japanese yen (JPY). Dental expenditure was excluded because it was unavailable for 37.8% of the primary cohort. Because costs were nonnegative, right-skewed, and included zeros, adjusted cost ratios were estimated using a generalized linear model with log link and Tweedie distribution. A power parameter of 1.7 was retained based on the lowest Akaike information criterion among candidate values evaluated. Expenditure analyses are interpreted as contemporaneous associations rather than prospective 365-day outcomes.

Ethics approval

The study was approved by the Research Ethics Committee of Kagawa Nutrition University (approval nos. 526 and 600). This study involved secondary analysis of de-identified administrative and health check-up data. No participants were prospectively recruited or contacted for this study, and individual informed consent was not obtained.

Data availability

The individual-level data are not publicly available because they were obtained under data-use agreements with participating municipalities and contain administrative and health check-up records subject to legal, ethical, and governance restrictions in Japan. De-identified data may be available upon reasonable request and subject to approval by the participating municipalities, relevant ethics committees, and data custodians where permitted.

Code availability

The statistical analysis code is not publicly deposited. Requests regarding the analytic code may be directed to the corresponding author and will be considered subject to applicable data-use, governance, and confidentiality restrictions.

Declarations

Acknowledgements

This work was supported by Health and Labor Sciences Research Grants (grant nos. 20AA2006 and 26AA2008) from the Ministry of Health, Labour and Welfare, Japan. The funder had no role in the study. The authors thank Editage for English language editing and manuscript review. OpenAI ChatGPT (GPT-5.6 Sol; accessed September 2026) was employed solely to support English language editing, verification of analysis code, and reorganization of manuscript text, preparation of draft tables and figures, and checking of analysis code. All statistical analyses were executed in SPSS. The authors directed, reviewed, and verified all AI-assisted output and take full responsibility for the accuracy and integrity of the manuscript. The AI tool was not used to generate or alter primary research data.

Author contributions

TT, YY, KT, and KI contributed to conceptualization, methodology, project administration, and validation and reviewed and edited the manuscript. TT, KT, and KI curated the data and contributed to investigation and resources. TT, YY, and KT conducted the formal analysis and wrote the original draft. TT and KT were responsible

for software. TT and KI supervised the study and contributed to visualization. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no financial or non-financial competing interests.

References

1. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. Lancet. 2019;394:1365–1375. doi:10.1016/S0140-6736(19)31786-6.
2. Venkatapuram S, Amuthavalli Thiyagarajan J. The Capability Approach and the WHO healthy ageing framework (for the UN Decade of Healthy Ageing). Age Ageing. 2023;52(suppl 4):iv6–iv9. doi:10.1093/ageing/afad126.
3. Tanaka T, Yoshizawa Y, Sugaya K, et al. Predictive validity of the Questionnaire for Medical Checkup of Old-Old for functional disability: using the National Health Insurance Database System. Geriatr Gerontol Int. 2023;23:124–130. doi:10.1111/ggi.14533.
4. Tanaka T, Yoshizawa Y, Kitamura K, et al. Frailty determined by the Questionnaire for Medical Checkup of Old-Old is correlated with increased healthcare cost: using the Japanese health insurance database system. Geriatr Gerontol Int. 2023;23:973–974. doi:10.1111/ggi.14711.
5. Castagna C, Huff A, Douglas A, et al. Stratifying the population based on health risk: identification of patient key health risk factors through consensus techniques. BMC Prim Care. 2025;26:229. doi:10.1186/s12875-025-02923-w.
6. Clegg A, Bates C, Young J, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing. 2016;45:353–360. doi:10.1093/ageing/afw039.
7. Ministry of Health, Labour and Welfare, Japan. Guidelines for Health Services Based on the Characteristics of Older Adults [Koreisha no Tokusei o Fumaeta Hoken Jigyo Guideline Dai 3-Han]. 3rd ed. Published March 29, 2024. Accessed September 25, 2026. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryouhoken/hokenjigyou/index_00003.html
8. Fukutomi E, Kimura Y, Wada T, Okumiya K, Matsubayashi K. Long-term care prevention project in Japan. Lancet. 2013;381:116. doi:10.1016/S0140-6736(13)60049-5.
9. Tsutsui T, Muramatsu N. Care-needs certification in the long-term care insurance system of Japan. J Am Geriatr Soc. 2005;53:522–527. doi:10.1111/j.1532-5415.2005.53175.x.
10. Kadu M, Ehrenberg N, Stein V, Tsiachristas A. Methodological quality of economic evaluations in integrated care: evidence from a systematic review. Int J Integr Care. 2019;19(3):17. doi:10.5334/ijic.4675.
11. Liljas AEM, Brattström F, Burström B, Schön P, Agerholm J. Impact of integrated care on patient-related outcomes among older people: a systematic review. Int J Integr Care. 2019;19(3):6. doi:10.5334/ijic.4632.
12. Satake S, Arai H. Questionnaire for medical checkup of old-old (QMCOO). Geriatr Gerontol Int. 2020;20:991–992. doi:10.1111/ggi.14004.
13. Li J, Yasuoka M, Kinoshita K, et al. Validity of the Questionnaire for Medical Checkup of Old-Old (QMCOO) in screening for physical frailty in Japanese older outpatients. Geriatr Gerontol Int. 2022;22:902–903. doi:10.1111/ggi.14472.

14. Yakabe M, Shibasaki K, Hosoi T, et al. Effectiveness of the Questionnaire for Medical Checkup of Old-Old (QMCOO) to diagnose frailty in community-dwelling older adults. Geriatr Gerontol Int. 2022;22:127–131. doi:10.1111/ggi.14328.
15. Matsuda S, Fujimori K. The claim database in Japan. Asian Pac J Dis Manag. 2014;6(3–4):55–59. doi:10.7223/apjdm.6.55.
16. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378.
17. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001;56:M146–M156. doi:10.1093/gerona/56.3.M146.
18. Kiuchi Y, Tsutsumimoto K, Nishimoto K, et al. Association of trajectories of physical frailty status and incident disability in community-dwelling older adults: a longitudinal study. Geriatr Gerontol Int. 2025;25:220–225. doi:10.1111/ggi.15063.
19. Tanaka T, Hirano H, Ikebe K, et al. Oral frailty five-item checklist to predict adverse health outcomes in community-dwelling older adults: a Kashiwa cohort study. Geriatr Gerontol Int. 2023;23:651–659. doi:10.1111/ggi.14634.
20. Tanaka T, Takahashi K, Hirano H, et al. Oral frailty as a risk factor for physical frailty and mortality in community-dwelling elderly. J Gerontol A Biol Sci Med Sci. 2018;73:1661–1667. doi:10.1093/gerona/glx225.
21. Diaz G, Prado CM, Gonzalez MC, Rugeles MP. Usefulness of the global leadership initiative on malnutrition - GLIM- criteria to identify malnutrition in older adults: systematic review and meta-analyses. Arch Gerontol Geriatr. 2026;142:106121. doi:10.1016/j.archger.2025.106121.
22. Garre-Olmo J, Calvó-Perxas L, López-Pousa S, de Gracia Blanco M, Vilalta-Franch J. Prevalence of frailty phenotypes and risk of mortality in a community-dwelling elderly cohort. Age Ageing. 2013;42:46–51. doi:10.1093/ageing/afs047.
23. Kojima T, Mizukami K, Tomita N, et al. Screening Tool for Older Persons' Appropriate Prescriptions for Japanese: report of the Japan Geriatrics Society Working Group on "Guidelines for medical treatment and its safety in the elderly." Geriatr Gerontol Int. 2016;16:983–1001. doi:10.1111/ggi.12890.
24. Tanaka T, Akishita M, Kojima T, Son BK, Iijima K. Polypharmacy with potentially inappropriate medications as a risk factor of new onset sarcopenia among community-dwelling Japanese older adults: a 9-year Kashiwa cohort study. BMC Geriatr. 2023;23:390. doi:10.1186/s12877-023-04012-y.
25. Tanaka T, Akishita M, Kojima T, Son BK, Iijima K. Anticholinergic burden quantified using the Japanese risk scale as a predictor of frailty and sarcopenia among community-dwelling older adults. Geriatr Gerontol Int. 2025;25:520–527. doi:10.1111/ggi.70012.
26. Kidney Disease: Improving Global Outcomes Diabetes Work Group. KDIGO 2022 clinical practice guideline for diabetes management in chronic kidney disease. Kidney Int. 2022;102:S1–S127. doi:10.1016/j.kint.2022.06.008.
27. Sum G, Choo RWM, Nai ZL, et al. A novel integrated geriatric services hub for frailty identification and comprehensive management of community-dwelling older adults in Singapore: impact on health service utilization. J Am Geriatr Soc. 2025;73:1576–1587. doi:10.1111/jgs.19339.
28. Kim DH, Schneeweiss S, Glynn RJ, Lipsitz LA, Rockwood K, Avorn J. Measuring frailty in Medicare data: development and validation of a claims-based frailty index. J Gerontol A Biol Sci Med Sci. 2018;73:980–987.

doi:10.1093/gerona/glx229.

29. Tanaka T, Lyu W, Yoshizawa Y, Son BK, Iijima K. Predictive validity of senior volunteer-led frailty check-up results for disability and mortality among community-dwelling older adults: a cohort study. Arch Gerontol Geriatr. 2025;139:105998. doi:10.1016/j.archger.2025.105998.
30. Tanaka T, Yoshizawa Y, Lyu W, Iijima K. Long-term impact of a senior volunteer-driven frailty check programme on care dependency and expenditures: a 10-year population-based cohort study in Japan. Age Ageing. 2026;55(6):afag161. doi:10.1093/ageing/afag161.

Figures

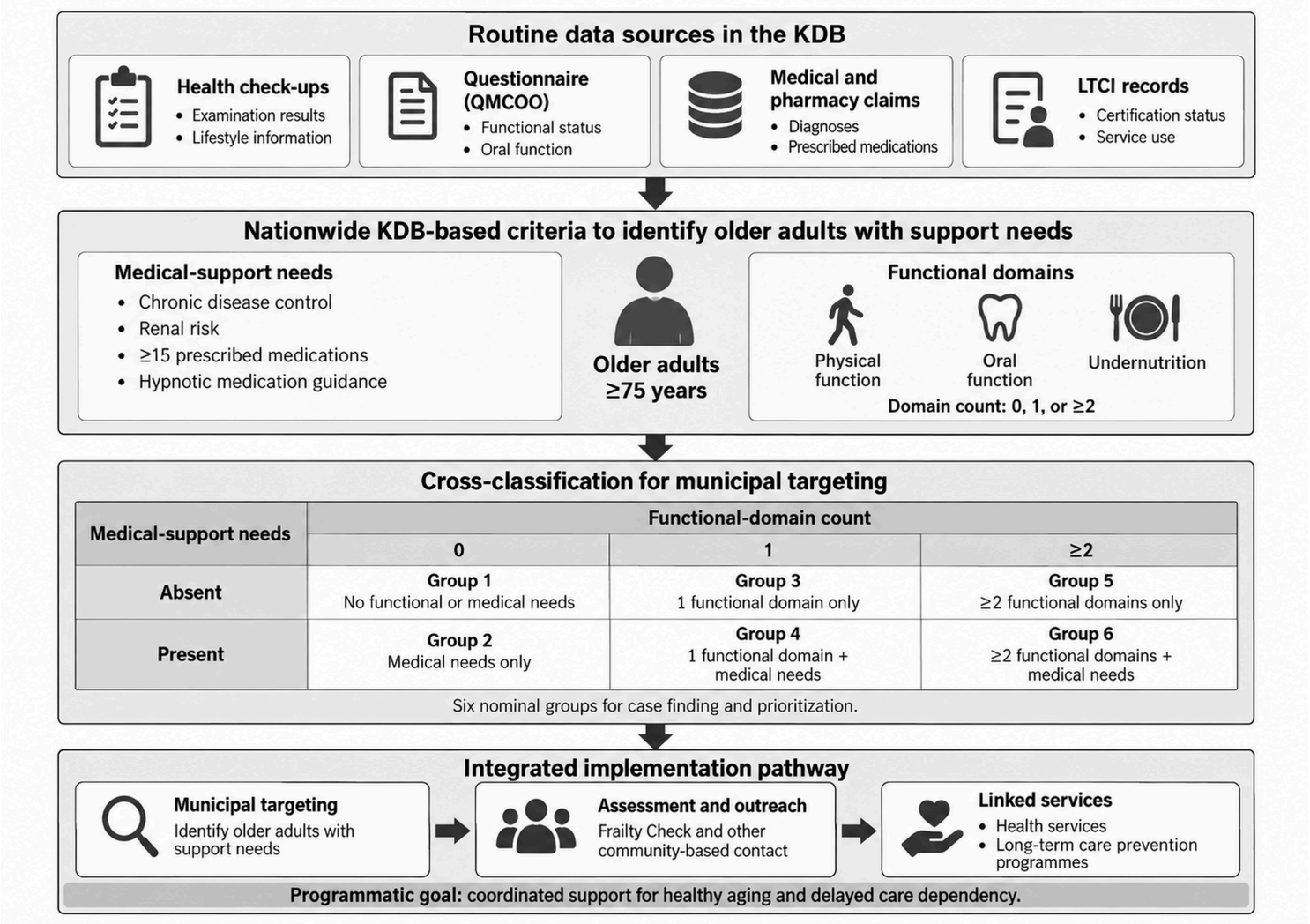


Figure 1

Routine-data framework for multidomain vulnerability stratification and care-system follow-up. Routinely collected health check-up/QMCOO, medical and pharmacy claims, and LTCI data feed the KDB-based framework. Functional-domain count (0, 1, or ≥2) and medical-support needs (absent or present) are cross-classified into six nominal groups used for case finding and prioritization. The present study evaluates 365-day LTCI certification, repeated-assessment transitions, and operating characteristics relevant to preventive follow-up. KDB, Kokuho Database; LTCI, long-term care insurance; QMCOO, Questionnaire for Medical Checkup of Old-Old.

Alt text: Flow diagram showing routinely collected health check-up/QMCOO, medical/pharmacy claims, and LTCI data feeding a KDB-based targeting tool. The tool cross-classifies functional vulnerability (0, 1, or ≥2 domains) with medical-support needs (absent/present) into six nominal groups, which are evaluated for 365-day LTCI certification, repeated-assessment transitions, and operational prioritization. The six groups are not an ordinal score.

Supplementary Files

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

- NPJAgingFrailtySupplementaryInformation.docx