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.

Joint capacity–deficit states, transitions and disability in older Chinese adults: a longitudinal cohort study

Fan Wang
Ningxia Medical University
Mengting Dai
Ningxia Medical University
Huili Yang
Ningxia Medical University
Meng Cao
Ningxia Medical University
Yanjin Huang
Ningxia Medical University
Haijiao Zhang
People's Hospital of Ningxia Hui Autonomous Region
Minhui Liu
mliu62@jhu.edu
Ningxia Medical University

Research Article

Keywords: intrinsic capacity, deficit burden, disability, activities of daily living, mortality, longitudinal cohort, healthy aging

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11509602/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

Background

Joint capacity–deficit states may describe different patterns of vulnerability, but incomplete measurements may affect estimated transitions.

Methods

CHARLS participants aged ≥ 60 years (2011–2015) were classified using a four-domain non-psychological capacity proxy and a 19-item deficit index: higher capacity/lower deficits (S1), lower capacity/lower deficits (S2), higher capacity/higher deficits (S3), and lower capacity/higher deficits (S4). Analyses used sampling/selection weights. Post hoc sensitivities recovered logically identifiable categories without imputing missing domains and refitted nuisance models within 200 stratified-cluster bootstrap samples.

Results

Among 4,932 positive-weight complete-origin/destination records, S2-to-S1 and S2-to-S4 proportions were 37.21% and 17.58%; S3-to-S1 and S3-to-S4 were 13.89% and 31.85%. Logical classification added 5,787 classifiable person-waves, leaving 24.7% unclassified, and yielded 8,604 positive-weight transitions. S2-to-S1 and S2-to-S4 became 25.46% and 27.48%, whereas S3-to-S4 remained more frequent than S3-to-S1 (34.01% versus 12.98%). Original survivor-conditional disability probabilities ranged from 11.94% in S1 to 33.08% in S4. Direct discordant-state outcome contrasts were inconclusive after multiplicity adjustment; the death-or-disability analysis retained 151 deaths and 960 disability events.

Conclusions

Joint states describe functional vulnerability, but the apparent S2 preference for S1 did not persist when logically identifiable incomplete records were included. Measurement definitions, residual missingness and absent external validation limit interpretation.

Background

The World Health Organization defines intrinsic capacity as the composite of an individual's physical and mental capacities and places it within a function-centered model of healthy aging [1, 2]. Intrinsic capacity has been operationalized and evaluated in a longitudinal Chinese cohort [3]; this does not validate the modified proxy used here. Frailty and deficit accumulation instead describe vulnerability arising from declining reserve and accumulated health problems [4]. Recent conceptual perspectives argue that intrinsic capacity and frailty should be integrated rather than treated as competing labels: capacity may identify functional resources and earlier decline, whereas frailty identifies clinically relevant vulnerability [5, 6].

A Dutch cohort study described age-related overlap and discordance using a five-domain intrinsic-capacity measure and frailty [7]. Four-group combinations have also been related to self-rated health in a cross-sectional study [8]. Longitudinal work has examined correlated changes [9], transitions in the two constructs in relation to disability [10], and five-domain capacity transitions as predictors of Fried-phenotype frailty transitions in a Korean cohort [11]. These studies establish substantial prior work; the present analysis does not introduce joint classification itself. The remaining question is not whether the constructs are related, but how people move between discordant joint states and how those states relate to subsequent functional outcomes when direct item overlap is reduced.

We therefore cross-classified a four-domain non-psychological intrinsic-capacity proxy with a reduced-overlap deficit index in the China Health and Retirement Longitudinal Study (CHARLS). Our primary objective was to describe transitions among four joint states. Secondary analyses examined associations with mortality and incident activities-of-daily-living (ADL) disability. The contribution is an explicit joint-state transition description and an assessment of its dependence on measurement definitions, rather than a new frailty scale or prediction tool. Additional analyses and the prioritization of survivor-conditional disability were post hoc, as detailed in the Methods.

Methods

Study design and participants

We used the 2011, 2013 and 2015 waves of Harmonized CHARLS Version D (June 2021; H_CHARLS_D_Data.dta), developed by the Gateway to Global Aging Data [12]. Detailed supporting methods are provided in Additional file 1 (eMethods 1–13). The STROBE checklist accompanies this submission. CHARLS is a nationally representative longitudinal survey of Chinese adults aged 45 years or older and their spouses [13]. The eligible frame comprised person-waves from respondents aged 60 years or older who completed an individual interview. Study size was determined by the eligible records available in the selected waves; sample flow, origin-record characteristics and phenotype missingness are documented in eTables 1–3. This secondary observational analysis used two approximately two-year intervals. Participants with an observed origin phenotype contributed person-period records for 2011–13 and 2013–15. Exposures and covariates were measured at the beginning of each interval and outcomes at the subsequent wave. In the harmonized dataset used for this analysis, the 2018 body mass index, grip strength and self-rated health fields could not reproduce the same phenotype definitions. We did not reconstruct these indicators from the raw 2018 files; the restriction to 2011–2015 therefore reflects the present harmonization, not evidence that comparable later-wave measurement is impossible.

The original CHARLS study received ethics approval from Peking University (IRB00001052-11015), and participants provided informed consent. The present secondary analysis of de-identified data was approved by the Ethics Committee of Ningxia Medical University (Review Opinion No. 2025-N263).

Joint capacity–deficit states

The modified intrinsic-capacity measure was explicitly a four-domain non-psychological proxy, not the complete five-domain WHO construct. It comprised cognition, locomotion, vitality and sensory capacity. Cognition combined immediate and delayed recall, serial subtraction, temporal orientation and figure copying. Impairment was defined as more than one standard deviation below the wave-specific mean among participants aged at least 60 years. Locomotor impairment was difficulty walking 1 km or rising from a chair. Vitality impairment was body mass index < 18.5 kg/m² or low grip strength (< 28 kg in men or < 18 kg in women), using the Asian Working Group cut-offs for low muscle strength [14]. These cut-offs were not developed to validate our combined vitality-domain definition. Sensory impairment was poor distant vision, near vision or hearing. Lower capacity required at least two impaired domains.

The 19-item reduced-overlap deficit index excluded direct capacity-domain, depressive-symptom and ADL/IADL items. It comprised 12 chronic conditions, five selected physical-function difficulties, hospitalization and self-rated health. It was calculated when at least 16 items were observed; higher burden was defined as ≥ 0.25. We use ‘deficit burden’, rather than clinical frailty, for this study-specific index. Cross-classification produced S1 (higher capacity/lower deficits), S2 (lower capacity/lower deficits), S3 (higher capacity/higher deficits) and S4 (lower capacity/higher deficits). Detailed definitions are in eMethods 1.

Outcomes and covariates

The main outcome was the subsequent four-state phenotype. Mortality before the next wave was examined as a mutually exclusive fifth destination. Incident ADL disability was any new difficulty in dressing, bathing, eating, transferring, toileting or continence among records with no baseline ADL difficulty and an observed subsequent ADL assessment.

Prognostic-model covariates were age (years), sex, education, partnered status, rural residence, ever smoking, past-year alcohol use and interval. Age and the harmonized education code (1–10) entered as linear terms; the latter assumes equal log-odds increments between codes rather than years of schooling. Sex and other binary predictors used 0/1 coding, and interval was categorical. Source variables and reference categories are specified in eTable 4. Because the 12-condition count directly contributed to the deficit index, its addition was treated as a sensitivity adjustment rather than part of the primary model. Depressive symptoms were measured with the 10-item Center for Epidemiologic Studies Depression Scale [15], and social participation indicated any listed activity during the previous month. Previous work examined participation in relation to intrinsic-capacity progression [16], frailty [17, 18], and depressive symptoms [19]. These studies used different measures and analytic frameworks; our corresponding exploratory analyses are confined to Additional file 1.

Statistical analysis

We reported sampling-weighted state distributions and selection-weighted transition proportions. Logistic observation models included demographic, lifestyle, chronic-condition, CES-D, participation, wave/interval and original predictor-missingness variables; follow-up and outcome-observation models also included origin joint state. Missing predictors were replaced by wave-specific medians, with pooled medians as fallback. This was single imputation; the primary Taylor intervals did not propagate uncertainty from estimated medians or weights. Original adjustment-covariate missingness was minimal (eTable 5), unlike phenotype missingness. Outcome-regression-unadjusted estimates are retained in eTable 6. Stabilized inverse-probability weights were truncated at the 1st and 99th percentiles, multiplied by interval-normalized sampling weights and truncated again. Taylor linearization used public-ID county/district groups as analysis pseudo-strata and villages/communities as clusters. The variance universe retained all 449 communities in the age-eligible frame, including zero-score communities outside each analysis subset (150 pseudo-strata; 299 degrees of freedom; no singleton strata). This approximation does not reconstruct every stage of the original systematic probability-proportional-to-size sampling design (eMethods 13). Counts and weighted percentages have explicitly labelled denominators. Original weighting diagnostics, missing-destination bounds, interval-specific results and five-destination estimates are in eTables 7–11.

The primary research objective remained four-state transitions. Mortality before the next wave was a separate destination and was also modeled among records with known vital status. For disability, the principal revised estimand was an association among baseline ADL-free records known alive at follow-up. Deaths and unknown vital status were excluded before fitting ADL-observation weights; survivors without ADL assessment remained nonrespondents. This is a survivor-conditional description, not a causal effect among people who would survive under alternative states. A parallel analysis classified baseline ADL-free records as alive without disability, alive with disability, or dead, and modeled death or disability versus alive without disability (eTables 12–14). Known survivors without ADL and records with unknown vital status were not coded as event-free. Neither endpoint identifies transient disability, disability preceding death or time-to-first-event cumulative incidence.

Weighted logistic outcome models adjusted for age, sex, education, partnered status, rural residence, ever smoking, past-year alcohol use and interval. Standardized probabilities averaged predictions under each state over the respective weighted analytic covariate distribution; delta-method confidence intervals used the Taylor covariance, conditional on estimated weights and the observed covariate distribution. These are model-standardized descriptive associations, not intervention effects. The original ADL model, whose observation weights treated deaths as absent ADL assessments, was retained only as a legacy sensitivity analysis. Its previously computed 200-replicate participant-cluster bootstrap performance estimates were also retained as exploratory results, not validation of the newly prioritized survivor model. That bootstrap held imputed values and estimated weights fixed and did not reproduce the stratified-PSU design. Earlier categorical-education, singleton-variance, original prognostic-model and validation analyses remain in eTables 15–18. Complete-covariate and chronic-count adjustments and weight-truncation checks are retained in eTables 19–21; chronic conditions are constituent exposure items, so adjustment changes the comparison rather than definitively controlling confounding.

Earlier measurement checks used a fixed 2011 cognitive threshold, all 19 deficit items, and a conventional 34-item frailty index using deficit-accumulation principles [20]; that source does not validate the present 19-item index. A depressive-symptom proxy was added as a fifth domain using CES-D-10 ≥ 12 [21]. The cited threshold study involved older Chinese adults referred for psychiatric assessment; its cutoff does not establish diagnostic validity in CHARLS or capture the full psychological-capacity domain. The fixed-threshold, item-completeness and five-domain transition results are in eTables 22–23; the earlier 34-item and other threshold checks concerned exploratory interactions, not validation of the present transition and disability results. Model-dependent depressive-symptom-by-participation findings remain in eTable 24 and eFigure 2. The original direct outcome contrasts are preserved in eTable 25.

New post hoc analyses directly addressed the revised main questions (eMethods 11; eTables 26–34). For each outcome we assessed observation rates, probability bounds, weight truncation, effective sample size and covariate balance within origin-state-by-interval cells against the baseline-weighted eligible sample. Flexible age and categorical education terms were tested in both the outcome-observation and outcome-regression models, retaining the upstream phenotype weights. We recalculated full transition matrices and survivor-conditional disability associations at deficit-index thresholds of 0.20 and 0.30 and a capacity threshold of at least one impaired domain, refitting selection weights and reporting reclassification. Continuous deficit burden per 0.10 and the number of impaired capacity domains were modeled jointly, with an exploratory interaction. These specifications change the operational exposure and do not test identical populations within each state. A further post hoc sensitivity omitted only missingness-indicator covariates flagged in fewer than 10 eligible records from endpoint-observation models, retaining all records, underlying predictors and upstream phenotype weights (eTable 35).

Earlier singleton-contribution checks are retained as historical sensitivities. Current transition and outcome inference uses the full age-eligible community universe; alternative community and county clustering specifications were examined without selecting by

significance (eMethods 13; eTables 40–42). Within-origin contrasts compared S2-to-S1 versus S2-to-S4 and S3-to-S4 versus S3-to-S1. These descriptive contrasts are distinct from comparing S2 and S3 outcome odds. The revised four-test Holm family comprised direct S3-versus-S2 contrasts for mortality, survivor ADL, death or disability, and survivor ADL additionally adjusted for chronic-condition count. All confidence intervals are unadjusted 95% intervals; other new sensitivity analyses were exploratory without multiplicity adjustment. Estimated weights remained fixed for Taylor inference. The primary specification was not selected according to significance or discrimination.

Two additional post hoc analyses addressed incomplete classification and nuisance-estimation uncertainty (eMethods 12; eTables 36–39). With k known impaired domains and m unknown domains, lower capacity was logically identifiable if $k \geq 2$ and higher capacity if $k + m < 2$. We retained the deficit-index observation rule, did not impute domains and refitted selection weights in this expanded sample. Separately, 200 ordinary within-stratum PSU bootstrap samples re-estimated predictor medians, both observation-weight stages, truncation and outcome regressions under the original complete-domain specification. Resampling used all 449 community clusters in 150 eligible-frame pseudo-strata, none of which was a singleton. Measurement definitions and supplied survey weights remained fixed. Exploratory percentile intervals therefore address some nuisance-estimation uncertainty, not the full sampling/measurement/missing-data process. They neither validate a prediction tool nor replace the primary intervals.

Analysis used Python 3.12.7, pandas 3.0.5 and statsmodels 0.15.0. Original internal validation used seed 20260913; the subsequent sensitivity analyses used seed 20260929. No independent external validation or full-design, raw-measurement reconstruction bootstrap was performed.

AI-assisted tools

OpenAI ChatGPT and Codex were used under author supervision for code drafting and debugging, language editing and document preparation. The authors verified the code, numerical results and manuscript and retain full responsibility for the work.

Results

Participants and missingness

The three waves contained 25,658 eligible person-waves from 11,619 participants. The joint phenotype was observed for 13,530 records (52.7%) from 7,785 participants. Missing cognition (28.1–34.1% by wave) and vitality (23.0–30.6%) contributed most to phenotype incompleteness. Among 8,149 records with an observed origin state, 5,002 had an observed subsequent state, 230 ended in death, 2,706 were known alive without a complete subsequent phenotype and 211 had unresolved subsequent status (eFigure 1).

The observed-transition analysis included 4,932 positive-weight records from 3,374 participants; 70 of the 5,002 observed transitions lacked a positive sampling weight. The combined weight had median 0.667, maximum 3.500 after truncation and effective sample size 3,453 (70.0% of the weighted analytic count). Post-weighting absolute standardized differences for assessed predictors were < 0.10 (eMethods 2).

At 2011 cohort baseline, 4,011 participants had an observed phenotype: 2,120 in S1, 740 in S2, 484 in S3 and 667 in S4. Mean depressive-symptom scores were highest in S4, whereas mean chronic-condition burden was highest in S3 (Table 1). The pooled 8,149 analytic origin records are described separately in eTable 2.

Table 1
Baseline profile by joint capacity–deficit state, 2011

Characteristic	Overall	S1	S2	S3	S4
Age, years, mean (SD)	67.60 (6.24)	66.55 (5.53)	69.26 (7.05)	66.82 (5.45)	69.85 (7.06)
Women, %	43.8	37.5	49.4	49.6	53.6
Chronic conditions, mean (SD)	1.66 (1.45)	1.06 (0.94)	0.99 (0.87)	3.42 (1.32)	2.87 (1.47)
Depressive-symptom score, mean (SD)	8.13 (6.05)	5.93 (4.80)	9.65 (6.15)	9.73 (5.91)	12.26 (6.54)
Any social participation, %	49.5	51.0	47.4	55.7	42.0
Unweighted n	4011	2120	740	484	667

Note. Values are survey-weighted except sample sizes. SD, standard deviation. S1, higher capacity/lower deficits; S2, lower capacity/lower deficits; S3, higher capacity/higher deficits; S4, lower capacity/higher deficits.

Weighted transitions

S1 and S4 were the most persistent states: 69.11 (66.90–71.23) and 61.47 (57.52–65.27), respectively, remained unchanged. Among S2 records, 37.21 (32.86–41.77) moved to S1, 34.89 (30.72–39.30) remained in S2 and 17.58 (14.63–20.98) moved to S4. Among S3 records, 13.89 (10.77–17.75) moved to S1, 49.97 (45.53–54.41) remained in S3 and 31.85 (27.69–36.32) moved to S4. All estimates are percentages with 95% confidence intervals (Table 2 and Fig. 1).

Table 2
Pooled weighted transitions between joint capacity–deficit states

Origin	To S1, n; % (95% CI)	To S2, n; % (95% CI)	To S3, n; % (95% CI)	To S4, n; % (95% CI)
S1	2023; 69.11 (66.90–71.23)	335; 13.53 (11.98–15.24)	286; 10.58 (9.26–12.05)	175; 6.79 (5.75–7.99)
S2	278; 37.21 (32.86–41.77)	243; 34.89 (30.72–39.30)	77; 10.32 (7.89–13.40)	129; 17.58 (14.63–20.98)
S3	97; 13.89 (10.77–17.75)	26; 4.28 (2.60–6.99)	378; 49.97 (45.53–54.41)	209; 31.85 (27.69–36.32)
S4	60; 6.94 (5.21–9.19)	52; 8.34 (6.16–11.20)	200; 23.25 (20.13–26.70)	434; 61.47 (57.52–65.27)

Note. S1, higher capacity/lower deficits; S2, lower capacity/lower deficits; S3, higher capacity/higher deficits; S4, lower capacity/higher deficits; CI, confidence interval. Counts are unweighted among 5,002 records with an observed destination. Percentages and confidence intervals use 4,932 records with positive combined selection and sampling weights and eligible-frame Taylor linearisation (449 communities; 150 analysis pseudo-strata; 299 degrees of freedom).

Interval-specific estimates suggested that S2-to-S1 movement was more frequent in 2011–13 than 2013–15 (42.86% versus 32.89%), whereas S2-to-S4 movement was less frequent (13.82% versus 20.46%). These descriptive interval differences were not specified as formal contrasts (eTable 10).

Mortality, survivor-conditional disability and joint outcomes

The mortality model included 7,834 positive-weight records and 228 deaths. S4 had higher estimated mortality odds than S1 (OR 2.32 (1.52–3.52)); estimates for S2 and S3 were 0.99 (0.63–1.54) and 1.25 (0.76–2.06). Parentheses give 95% confidence intervals. The five-destination analysis used a different observation model and included 5,160 records with either a phenotype or recorded death; its mortality percentages should not be equated with those from the full known-vital-status sample (eTable 11).

Of 6,489 baseline ADL-free origin records, 6,180 were known alive at follow-up and 5,916 had an ADL assessment. The principal survivor-conditional model used 5,842 positive-weight records from 4,381 participants, with 960 disability events. ORs versus S1 were 1.73 (1.36–2.20), 2.43 (1.86–3.17) and 3.73 (2.90–4.81) for S2–S4 (Table 3). Standardized probabilities were 11.94%, 18.89%, 24.52% and 33.08%, respectively; differences versus S1 were 6.95, 12.58 and 21.14 percentage points (Table 4). These probabilities are conditional on known survival, not full-baseline-population risks.

Table 3
Mortality, survivor-conditional disability and death-or-disability associations

Estimate	Mortality	ADL among known survivors	Death or disability
Events / records	228 / 7834	960 / 5842	1111 / 5993
S2 versus S1, OR (95% CI)	0.99 (0.63–1.54)	1.73 (1.36–2.20)	1.60 (1.28–2.00)
S3 versus S1, OR (95% CI)	1.25 (0.76–2.06)	2.43 (1.86–3.17)	2.25 (1.74–2.92)
S4 versus S1, OR (95% CI)	2.32 (1.52–3.52)	3.73 (2.90–4.81)	3.53 (2.79–4.46)

Note. Adjusted for age, sex, education, partnered status, rural residence, ever smoking, past-year alcohol use and interval. Models use endpoint-specific sampling/selection weights and fixed-weight eligible-frame Taylor covariance (299 degrees of freedom). Disability is conditional on known survival; the composite concerns interval-endpoint status, not time-to-first-event risk. The survivor model was prioritized during revision, not prespecified. Previous-model discrimination estimates remain in eTable 18. S1, higher capacity/lower deficits; S2, lower capacity/lower deficits; S3, higher capacity/higher deficits; S4, lower capacity/higher deficits. ADL, activities of daily living; OR, odds ratio; CI, confidence interval.

Table 4
Model-standardized ADL disability probability among known survivors by joint state

Joint state	Standardized probability, % (95% CI)	Difference versus S1, pp (95% CI)
S1	11.94 (10.54–13.34)	Reference
S2	18.89 (16.05–21.72)	6.95 (3.65–10.24)
S3	24.52 (20.15–28.88)	12.58 (8.08–17.07)
S4	33.08 (28.69–37.47)	21.14 (16.37–25.91)

Note. Standardization uses the weighted covariate distribution of 5,842 positive-weight records from 4,381 participants. Observation weights were estimated among 6,180 baseline ADL-free records known alive, excluding deaths and unknown vital status beforehand. Delta-method confidence intervals use fixed-weight eligible-frame Taylor covariance (299 degrees of freedom), conditional on the empirical covariate distribution. These are descriptive survivor-conditional probabilities, not baseline-population cumulative incidences or causal effects among always-survivors. pp, percentage points. S1, higher capacity/lower deficits; S2, lower capacity/lower deficits; S3, higher capacity/higher deficits; S4, lower capacity/higher deficits. ADL, activities of daily living; OR, odds ratio; CI, confidence interval.

The joint-endpoint analysis left 421 records unclassified: 264 known survivors without ADL and 157 with unresolved vital status. Among 5,993 positive-weight classified records from 4,490 participants, 4,882 were alive without disability, 960 alive with disability and 151 dead. ORs for death or disability versus S1 were 1.60 (1.28–2.00), 2.25 (1.74–2.92) and 3.53 (2.79–4.46). Standardized composite probabilities were 14.12%, 20.63%, 26.69% and 36.01% (eTables 12–13). Disability contributed 86.4% of composite events.

Direct comparisons and sensitivity of the main findings

The revised direct S3-versus-S2 ORs were 1.27 (0.70–2.33) for mortality, 1.40 (1.03–1.91) for survivor ADL and 1.41 (1.05–1.90) for death or disability. Holm-adjusted P values were 0.864, 0.095 and 0.094. Additional chronic-condition adjustment gave a survivor-ADL contrast of 1.00 (0.69–1.45) (adjusted P = 0.992). These comparisons neither establish reliable outcome separation nor equivalence between discordant states (eTable 29).

With the original state definitions, S2-to-S1 exceeded S2-to-S4 by 19.62 (13.21–26.03) percentage points, and S3-to-S4 exceeded S3-to-S1 by 17.96 (11.56–24.36) points, using the eligible-frame variance specification (eTable 40). These within-origin contrasts do not demonstrate that whole transition distributions differ between S2 and S3. Fixed cognitive thresholds and requiring complete deficit items retained the previously reported focal directions; adding the CES-D proxy changed some transition cells by about six points (archived eTables 22–23).

Changing the deficit threshold to 0.20 or 0.30 reclassified 14.07% and 11.65% of complete person-waves, respectively, while retaining the two focal transition directions. Lowering the capacity threshold to one impaired domain reclassified 34.72%; S2-to-S1 and S2-to-S4 became nearly equal (20.49% and 20.04%), while S3-to-S4 rose to 52.86%. Survivor-ADL point estimates remained ordered S1 < S2<S3 < S4 across these definitions, but this numerical ordering was not a test of all pairwise differences. Thus, broad disability associations persisted while transition magnitudes and the S2 directional contrast were definition-sensitive (eTables 28 and 34).

In a joint additive continuous-component model, survivor-ADL odds increased per additional impaired capacity domain (OR 1.27 (1.15–1.40)) and per 0.10 higher deficit index (1.54 (1.42–1.67)). The interaction OR was 1.02 (0.95–1.10) (unadjusted P = 0.533), providing no

clear evidence of departure from the specified additive log-odds model (eTable 30). These estimates do not validate a clinical score. The original ADL model's corrected AUC was 0.671 for both the binary additive-component and four-state models after rounding; those legacy performance results are not validation of the current survivor model (eTable 18).

Selection and modeling diagnostics

Endpoint effective sample sizes were 5,851 for mortality, 4,424 for survivor ADL and 4,494 for the joint endpoint. Maximum absolute within-state-by-interval covariate differences after weighting were 0.076, 0.116 and 0.138, respectively. Some balance worsened after weighting; uniform balance below 0.10 was not achieved for the disability endpoints. Sparse predictor patterns also generated near-zero observation probabilities (eTables 26–27). Omitting rare missingness indicators increased the minimum predicted observation probabilities to 0.662, 0.816 and 0.699 for mortality, survivor ADL and the joint endpoint, respectively; the largest absolute change in any state OR was below 0.001 (eTable 35). This sensitivity retained every eligible record and did not replace the principal specification. Flexible age/categorical education refits yielded similar state associations, but do not establish correct specification or remove unmeasured selection (eTable 32).

Among known survivors with observed origins, 38.7% of base-weighted records lacked a destination phenotype. Previously calculated reversal thresholds were attainable, so unmeasured selection could overturn focal transition directions (eTables 8–9). The additional origin-missingness profile describes measured differences but does not recover outcomes for missing origin states (eTable 33). Main adjustment-covariate missingness was minimal and should not be conflated with the 47.3% overall phenotype incompleteness.

Logical classification recovered 5,787 additional person-waves without imputing any missing domain: 19,317 of 25,658 records were classifiable and 6,341 (24.7%) remained unclassified. All originally observed states were unchanged. The expanded analysis included 11,910 origin records and 8,604 positive-weight transitions. S2-to-S1 was 25.46% and S2-to-S4 was 27.48%; their difference was − 2.02 percentage points (95% CI − 6.55 to 2.51). S3-to-S4 remained more frequent than S3-to-S1 (34.01% versus 12.98%). Thus, the original S2 directional finding did not persist after including logically identifiable incomplete records. Expanded survivor-ADL ORs versus S1 were 2.13, 2.52 and 4.94; the changed magnitudes do not establish removal of selection bias (eTables 36–38).

The nuisance-refit bootstrap completed 200 of 200 attempted samples. Under the original complete-domain specification, percentile intervals for the S2 and S3 within-origin differences were 13.74–24.79 and 12.95–22.57 percentage points, respectively (eTable 39). These intervals describe resampling variability within that specification and do not resolve the contrasting logical-classification result. Retaining zero-score communities and using alternative cluster specifications did not change these focal interpretations or produce multiplicity-adjusted evidence separating S2 and S3 outcomes (eTables 40–42).

Discussion

This study describes movement among joint capacity–deficit states and their associations with subsequent functional outcomes. Under the original operational definition, S2 more often moved to S1 than S4, whereas S3 more often moved to S4 than S1. Survivor-conditional disability probabilities increased from about 12% in S1 to 33% in S4. However, the S2 transition preference was attenuated under a one-domain threshold and did not persist when incomplete records with logically identifiable categories were included under the original two-domain threshold. The state labels therefore describe threshold-dependent classifications, not established biological stages.

The study extends joint cross-sectional classification and prior longitudinal analyses by explicitly describing discordant-state destinations using measures with reduced direct item overlap. Its incremental value is the distinction between relatively consistent functional-outcome associations and definition- and completeness-sensitive transition patterns. It does not establish that joint classification improves care or that one discordant state is intrinsically more hazardous. The direct S3-versus-S2 outcome contrasts remained inconclusive after multiplicity adjustment, and continuous-component analysis did not show clear interaction on the modeled log-odds scale. These findings favor using the four groups as a descriptive framework rather than as a new predictive instrument. Using ICOPE-based screening items and the Clinical Frailty Scale in Singapore, Zhou and colleagues reported interactions between capacity losses and frailty for instrumental ADL and EQ-5D-5L outcomes [22]. Their measures, outcomes and modeling scales differ from ours; our interaction analysis therefore does not directly test replication of their findings. Chronic-count adjustment nearly eliminated the direct disability contrast, but the count is part of the deficit exposure; this changes the comparison and is not proof that ordinary confounding explains the findings.

Separating survivor-conditional disability from the death-or-disability endpoint makes the target populations explicit. Neither identifies a causal effect among always-survivors, and the composite is dominated by disability events. Joint destinations cannot reconstruct

disability onset before death or transient disability between interviews. Original and revised models give similar numerical estimates, but similarity does not make their estimands interchangeable. Likewise, the previously reported internal validation pertains to the original weighted ADL model and cannot be transferred to the newly prioritized model.

Strengths include temporally ordered observations, transparent handling of repeated person-period records, cluster-aware variance approximation, outcome-specific observation weighting, direct contrasts, and checks addressing the current main questions rather than only the earlier exploratory interaction. The measurement checks provide an important boundary: changing the deficit threshold retained the focal directions, while broadening lower capacity substantially changed the classification and weakened the S2 direction. Logical classification additionally showed that requiring all four domains can alter the observed transition pattern even without changing the two-domain threshold. This is a substantive boundary, not a nuisance removed by sensitivity analysis.

Several limitations remain. The capacity measure excludes the psychological domain; CES-D sensitivity does not constitute full psychological-capacity measurement. The 19-item index is disease-heavy and study-specific, with 12 chronic conditions, and remaining functional content may correlate with the capacity domains despite exclusion of identical items and ADL/IADL. Neither construct is a newly validated clinical instrument. Threshold crossing may reflect measurement variation rather than reversible physiological loss. The continuous model retains linear log-odds assumptions and cannot settle all measurement questions.

Complete-domain phenotypes were unavailable for 47.3% of eligible person-waves. Logical identification reduced unclassifiable categories to 24.7% without recovering the missing measurements. This expansion remained selective: records near the two-domain threshold were more likely to remain unresolved. Observation weighting requires adequate positivity and measured predictors sufficient for selection adjustment. Residual imbalance limits confidence in those assumptions. Extreme predictions were sensitive to sparse missingness-indicator coding, although omitting those indicators barely changed state associations. This finding neither establishes positivity for missing origin phenotypes nor removes unmeasured selection. Destination bounds concern observed origins and do not identify transitions for missing origin phenotypes. The current Taylor intervals hold estimated weights fixed and use county/district pseudo-strata, not the original region/urban-rural/GDP sorting strata. County-cluster sensitivities do not recover unknown joint inclusion probabilities or original survey calibration. The new nuisance-refit bootstrap propagates some estimation variability, but ordinary n-PSU resampling is not a calibrated survey replicate-weight procedure, and its 200 replicates provide limited tail precision. Measurement-threshold estimation, latent missing-domain values and original survey-weight calibration remain outside that calculation. New analyses and their prioritization were post hoc; unadjusted sensitivity intervals are descriptive. Community-dwelling respondents may not represent institutionalized adults. The 2011–2015 waves support internally consistent measurement but not contemporary prevalence estimates, and later raw-wave harmonization and external validation were not undertaken. Observational associations do not identify treatment targets.

Conclusions

Joint capacity–deficit states provide a transparent description of subsequent functional vulnerability, with important dependence of transitions on operational thresholds and completeness requirements. Direct outcome separation of the discordant states and clinical benefit from the classification remain unestablished. Replication with contemporary data, standardized measures and explicit handling of disability and death is needed before clinical application.

Abbreviations

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

EQ-5D-5L
EuroQol five-dimension five-level instrument
IADL
instrumental activities of daily living
IC
intrinsic capacity
ICOPE
integrated care for older people
OR
odds ratio
PPS
probability proportional to size
PSU
primary sampling unit
SD
standard deviation
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology
WHO
World Health Organization.

Declarations

Ethics approval and consent to participate

The original CHARLS study received ethics approval from Peking University (IRB00001052-11015), and participants provided informed consent. The present secondary analysis of de-identified data was approved by the Ethics Committee of Ningxia Medical University (Review Opinion No. 2025-N263).

Clinical trail number

Not applicable.

Consent for publication

Not applicable. The manuscript contains no identifiable individual participant information.

Availability of data and materials

The CHARLS public-use data are available to registered users through the official CHARLS data-access service, subject to its application and data-use conditions [23]. This analysis used Harmonized CHARLS Version D (June 2021); documentation and the access route are provided in references [12, 24]. The authors cannot redistribute the source microdata. Aggregate results and measurement specifications are included in the article and Additional file 1. The analysis code has not been deposited in a public repository.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Ningxia Natural Science Foundation (grant number 2026AAC031333) to Fan Wang. The funder had no role in study design; data collection, analysis or interpretation; manuscript preparation; or the decision to submit the manuscript for publication.

Authors' contributions

FW: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft, Supervision, Project administration and Funding acquisition. MTD: Methodology, Formal analysis, Validation and Writing – review & editing. HLY: Methodology, Validation, Investigation and Writing – review & editing. MC: Data curation, Validation, Visualization and Writing – review & editing. YJH: Conceptualization, Methodology, Supervision and Writing – review & editing. HJZ: Investigation, Validation and Writing – review & editing. MHL: Conceptualization, Methodology, Supervision and Writing – review & editing. All authors contributed to interpretation of the findings, critically reviewed the manuscript, approved the final version for submission and agreed to be accountable for their contributions and for the integrity of the work.

Acknowledgements

The authors thank the CHARLS research team and all study participants. The analysis used Harmonized CHARLS Version D (June 2021), developed by the Gateway to Global Aging Data with support from the US National Institute on Aging (R01 AG030153, RC2 AG036619 and R03 AG043052). These awards supported the harmonized resource, not this study.

References

1. World Health Organization. World report on ageing and health. Geneva: World Health Organization; 2015.
2. World Health Organization. Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. Geneva: World Health Organization; 2017.
3. Beard JR, Si Y, Liu Z, Chenoweth L, Hanewald K. Intrinsic capacity: validation of a new WHO concept for healthy aging in a longitudinal Chinese study. J Gerontol Biol Sci Med Sci. 2022;77:94–100. 10.1093/gerona/glab226.
4. Belloni G, Cesari M. Frailty and intrinsic capacity: two distinct but related constructs. Front Med (Lausanne). 2019;6:133. 10.3389/fmed.2019.00133.
5. Cesari M, Canevelli M, Amuthavalli Thiyagarajan J, et al. Frailty and intrinsic capacity: integrating complementary concepts to promote healthy ageing and transformation of care. Age Ageing. 2026;55:afag138. 10.1093/ageing/afag138.
6. Cacciatore S. Intrinsic capacity will not replace frailty—and should not try. Age Ageing. 2026;55:afag137. 10.1093/ageing/afag137.
7. Qi Y, van Schoor NM, Schaap LA, Hoogendijk EO. Disentangling the overlap between frailty and intrinsic capacity in older adults. Exp Gerontol. 2026;213:113006. 10.1016/j.exger.2025.113006.
8. Shiratsuchi D, Makizako H, Akaida S, et al. Combinations of intrinsic capacity and frailty and their associations with self-rated health in community-dwelling older adults. Aging Clin Exp Res. 2025;37:275. 10.1007/s40520-025-03189-z.
9. Liu S, Kang L, Liu X, et al. Trajectory and correlation of intrinsic capacity and frailty in a Beijing elderly community. Front Med (Lausanne). 2021;8:751586. 10.3389/fmed.2021.751586.
10. Jia S, Zhao W, Ge M, et al. Associations between transitions of intrinsic capacity and frailty status, and 3-year disability. BMC Geriatr. 2023;23:96. 10.1186/s12877-023-03795-4.
11. Cho HJ, Jung H, Won CW, Kim M. Intrinsic capacity transitions as predictors of frailty transitions in community-dwelling older adults: findings from the Korean Frailty and Aging Cohort Study. Arch Gerontol Geriatr. 2025;128:105637. 10.1016/j.archger.2024.105637.
12. Phillips D, Green H, Petrosyan S, Shao K, Wilkens J, Lee J, Harmonized CHARLS, Documentation, Version D. (2011–2018). University of Southern California; June 2021. Accessed September 29, 2026. https://charls.charlsdata.com/Public/ashelf/public/uploads/document/harmonized_charls/application/Harmonized_CHARLS_D.pdf
13. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol. 2014;43:61–8. 10.1093/ije/dys203.
14. Chen LK, Woo J, Assantachai P, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc. 2020;21:300–e3072. 10.1016/j.jamda.2019.12.012.

15. Boey KW. Cross-validation of a short form of the CES-D in Chinese elderly. Int J Geriatr Psychiatry. 1999;14:608–17. 10.1002/.(SICI)1099-1166(199908)14:8<608::AID-GPS991>3.0.CO;2-Z.
16. Xie G, Focacci CN, Li J, Wang R, Chen G. Association of social participation with progression and reversion of intrinsic capacity in older adults: based on multistate model. J Nutr Health Aging. 2025;29:100719. 10.1016/j.jnha.2025.100719.
17. Fang B, Huang J, Zhao X, Liu H, Chen B, Zhang Q. Concurrent and lagged associations of social participation and frailty among older adults. Health Soc Care Community. 2022;30:e4812–20. 10.1111/hsc.13888.
18. Lin J, Wang K, Zhang Y, et al. Frailty index trajectories in Chinese older adults with diverse levels of social participation: findings from a national population-based longitudinal study. Aging Clin Exp Res. 2023;35:3105–14. 10.1007/s40520-023-02617-2.
19. Shiba K, Torres JM, Daoud A, et al. Estimating the impact of sustained social participation on depressive symptoms in older adults. Epidemiology. 2021;32:886–95. 10.1097/EDE.0000000000001395.
20. Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr. 2008;8:24. 10.1186/1471-2318-8-24.
21. Cheng ST, Chan ACM. The Center for Epidemiologic Studies Depression Scale in older Chinese: thresholds for long and short forms. Int J Geriatr Psychiatry. 2005;20:465–70. 10.1002/gps.1314.
22. Zhou K, Ng YS, Tay EL, Mah SM, Tay L. Effects of intrinsic capacity on health-related outcomes are moderated by frailty status among community-dwelling older adults. J Nutr Health Aging. 2026;30(9):100936. 10.1016/j.jnha.2026.100936.
23. China Health and Retirement Longitudinal Study. CHARLS. https://charls.pku.edu.cn/en/. Accessed 26 Oct 2026.
24. China Health and Retirement Longitudinal Study. Harmonized CHARLS. https://charls.charlsdata.com/pages/Data/harmonized_charls/en.html. Accessed 26 Oct 2026.

Figures

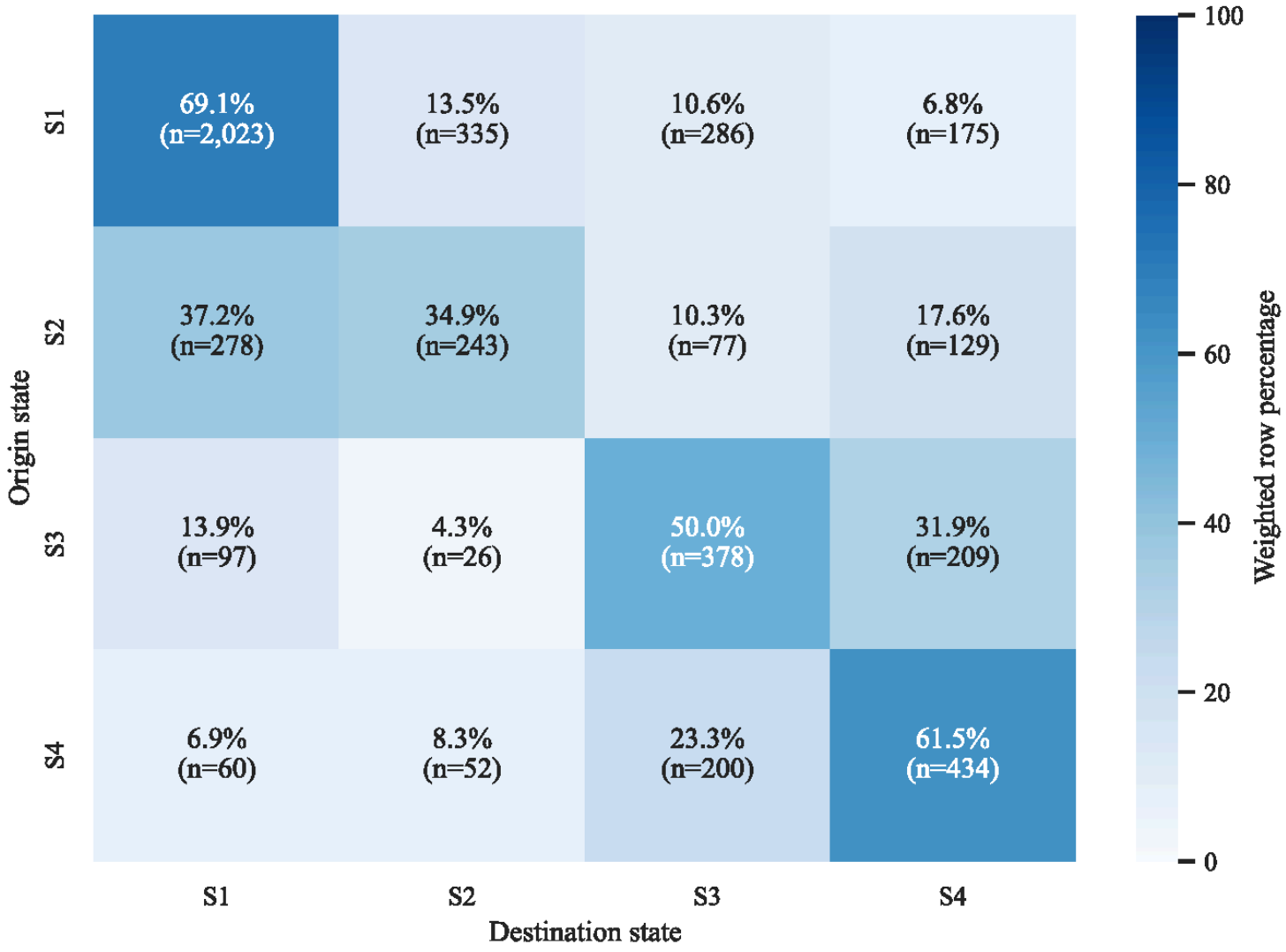


Figure 1

Weighted transitions between joint capacity–deficit states. Cells show weighted row percentages and unweighted cell counts. Counts include 5,002 records with an observed destination; weighted percentages use the 4,932 records with positive combined selection and sampling weights. Therefore, percentages cannot be reconstructed directly from the displayed counts. Confidence intervals in Table 2 use eligible-frame Taylor linearization with county/district analysis pseudo-strata. S1, higher capacity/lower deficits; S2, lower capacity/lower deficits; S3, higher capacity/higher deficits; S4, lower capacity/higher deficits.

Alt text: A four-by-four blue heat map shows that S1 and S4 are the most persistent states, S2 frequently moves to S1, and S3 frequently remains S3 or moves to S4.

Supplementary Files

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

- Additionalfile1.pdf