Additional file 1: Supplementary methods, tables and figures

eMethods 1: Detailed phenotype and outcome definitions
eMethods 2: Participant flow, missingness and weighting
eMethods 3: Interval-specific weighted transitions
eMethods 4: Mortality as a fifth destination
eMethods 5: Prognostic associations and model performance
eMethods 6: Threshold and item-completeness sensitivity analyses
eMethods 7: Exploratory depressive-symptom and social-participation analyses
eMethods 8: Additional construct and subgroup analyses
eMethods 9: Complex-survey and reproducibility notes
eMethods 10: Revision joint-endpoint sensitivity analyses
eMethods 11: Scientific-revision analyses and estimand prioritization
eMethods 12: Logical category identification and nuisance-refit resampling
eMethods 13: Eligible-frame variance integration and design sensitivity
eTable 1. Analytic flow
eTable 2. Pooled analytic origin-record profile by joint state
eTable 3. Missingness by wave
eTable 4. Source variables and analysis coding
eTable 5. Original missingness of primary adjustment covariates
eTable 6. Outcome-regression-unadjusted joint-state associations
eTable 7. Weight diagnostics
eTable 8. Worst-case bounds for focal transitions among records known alive
eTable 9. Reversal thresholds for the focal discordant-state orderings
eTable 10. Selected discordant-state transitions by interval
eTable 11. Weighted five-state destinations
eTable 12. Three-destination death–disability distribution
eTable 13. Adjusted death-or-disability composite
eTable 14. ADL disability among known survivors
eTable 15. Categorical-education composite sensitivity
eTable 16. Singleton-stratum variance sensitivity
eTable 17. Adjusted joint-state associations
eTable 18. Participant-cluster bootstrap internal validation
eTable 19. Complete-covariate sensitivity of fully adjusted joint-state associations
eTable 20. Sensitivity to additional adjustment for chronic-condition count
eTable 21. Sensitivity to ADL observation-weight truncation
eTable 22. Core weighted transition estimates under alternative definitions
eTable 23. Five-domain sensitivity using depressive symptoms as a psychological proxy
eTable 24. Exploratory interaction models
eTable 25. Original-model direct S3-versus-S2 outcome contrasts
eTable 26. Endpoint-specific weight and positivity diagnostics
eTable 27. State-by-interval ascertainment and balance
eTable 28. Definition sensitivity of focal transitions and disability associations
eTable 29. Revised direct discordant-state outcome contrasts
eTable 30. Continuous-component survivor-ADL models
eTable 31. Transition singleton correction and within-origin probability contrasts
eTable 32. Flexible age and education in observation and outcome models
eTable 33. Measured profile by origin-phenotype availability
eTable 34. Full transition matrices under alternative definitions
eTable 35. Sparse missingness-indicator sensitivity
eTable 36. Logical identification of categories in incomplete records
eTable 37. Transitions after including logically identified categories
eTable 38. Outcome associations after logical identification
eTable 39. Original-specification nuisance-refit bootstrap sensitivity
eTable 40. Focal transition contrasts under alternative variance specifications
eTable 41. Direct discordant-state outcome contrasts under alternative variance specifications
eTable 42. State associations under alternative clustering
eFigure 1. Participant and person-period flow.
eFigure 2. Exploratory quadratic depressive-symptom-by-social-participation model for S2-to-S4 progression.
Supplemental References

eMethods 1: Detailed phenotype and outcome definitions

Four-domain non-psychological intrinsic-capacity proxy

The measure comprised cognition, locomotion, vitality and sensory capacity. It omitted the psychological domain because depressive symptoms were examined separately and therefore must not be interpreted as the complete World Health Organization five-domain construct (1).

Cognition combined the mean of immediate and delayed recall (0–10), serial subtraction (0–5), temporal orientation (0–4) and figure copying (0–1), giving a total of 0–20. The main definition classified impairment below the wave-specific mean minus one standard deviation among participants aged at least 60 years. A sensitivity analysis applied the fixed 2011 threshold of 7.07 throughout. Locomotor impairment was difficulty walking 1 km or rising from a chair. Vitality impairment was body mass index <18.5 kg/m² or grip strength <28 kg in men and <18 kg in women (2). Sensory impairment was poor distant vision, near vision or hearing. All four domains had to be observed, and lower capacity required at least two impaired domains.

Reduced-overlap deficit index

The 19-item index comprised hypertension, dyslipidaemia, diabetes, cancer, chronic lung disease, liver disease, heart disease, stroke, kidney disease, digestive disease, arthritis and asthma; difficulty climbing stairs, stooping/crouching, lifting approximately 5 kg, picking up a small coin and raising the arms; hospitalisation in the previous year; and self-rated health. Binary items were scored 0/1; self-rated health used the harmonized r1shlt, r2shlt and r3shlt variables, with excellent/very good (codes 1/2) scored 0, good (3) scored 0.5, and fair/poor (4/5) scored 1. The analysis did not merge the alternative r*shlta responses into these variables; any harmonization of the original questions was inherited from the supplied harmonized file. The index was the mean when at least 16 of 19 items (≥80%) were observed. Higher deficit burden was ≥0.25. Direct intrinsic-capacity, depressive-symptom and ADL/IADL items were excluded to reduce construct overlap. Because the index was intentionally abbreviated and modified, the manuscript does not label it a clinical frailty diagnosis.

S1 denoted higher capacity/lower deficits; S2, lower capacity/lower deficits; S3, higher capacity/higher deficits; and S4, lower capacity/higher deficits.

Patient-important outcomes

Death before the subsequent wave was identified from CHARLS interview-status variables. Incident ADL disability was any new reported difficulty in dressing, bathing, eating, transferring, toileting or controlling urination/defecation among records with zero baseline difficulties and an observed subsequent ADL assessment.

eMethods 2: Participant flow, missingness and weighting

Supporting results are provided in eTables 1, 2, 3.

Weight construction and diagnostics

Phenotype-observation and follow-up-observation probabilities were estimated with logistic models containing age, sex, education, partnered status, rural residence, ever smoking, past-year alcohol use, chronic-condition count, CES-D score, social participation, wave/interval and predictor-missingness indicators. Follow-up observation models also included origin joint state, as confirmed in the analysis code. Missing continuous and binary predictor values were replaced by the median within wave, with the pooled median used only if a wave-specific median was unavailable; indicators retained whether each value was originally missing. The stabilising numerator was the wave-specific phenotype-observation rate for phenotype weights and the state-by-interval follow-up-observation rate for follow-up weights. Stabilised inverse-probability weights were truncated at the 1st and 99th percentiles at each stage. The final weight multiplied phenotype and follow-up weights by the sampling weight normalised within interval and was truncated again at the 1st and 99th percentiles. Of 5,002 observed transitions, 4,932 had a positive combined weight.

Supporting results are provided in eTables 7, 8, 9.

eMethods 3: Interval-specific weighted transitions

Supporting results are provided in eTable 10.

eMethods 4: Mortality as a fifth destination

A separate observation weight was estimated for records with either an observed subsequent phenotype or recorded death. The positive-weight sample contained 5,160 records (effective sample size 3,635); 1st and 99th percentile truncation limits were 0.251 and 3.541.

Supporting results are provided in eTable 11.

eMethods 5: Prognostic associations and model performance

Supporting results are provided in eTables 17, 18, 20, 21, 19.

eMethods 6: Threshold and item-completeness sensitivity analyses

Supporting results are provided in eTables 22, 23.

eMethods 7: Exploratory depressive-symptom and social-participation analyses

Linear target-transition models

The weighted logistic model contrasted each target transition with all other observed destinations from the same origin state. This binary estimand answers whether the probability of a selected destination differs from the combined probability of all alternatives; it does not distinguish persistence from movement to other states.

The models adjusted for age, sex, education, partnered status, rural residence, ever smoking, alcohol use, chronic-condition count and interval. CES-D was entered per five points without centering. Consequently, the CES-D main-effect odds ratio applies to the non-participation group, the social-participation main-effect odds ratio applies at CES-D=0, and the interaction odds ratio is the ratio of the per-five-point CES-D odds ratios between participation groups. These conditional coefficients should not be read as average effects across the CES-D distribution. Supporting results are provided in eTable 24.

eMethods 8: Additional construct and subgroup analyses

A conventional 34-item frailty index included 12 chronic conditions, six ADL items, five IADL items, nine physical-function items, hospitalisation and self-rated health. It was calculable for 95.17% of eligible person-waves and correlated 0.913 with the reduced-overlap index and 0.544 with the number of impaired capacity domains. The S2-to-S4 exploratory interaction odds ratios were 1.52 (1.09–2.12) at a 0.25 threshold and 1.78 (1.16–2.72) at 0.20. Subgroup analyses by sex, residence and multimorbidity were exploratory. No three-way interaction met a Benjamini–Hochberg false-discovery-rate threshold of $q<0.05$; the smallest q value was 0.079. Subgroup point estimates are therefore not interpreted as evidence of effect heterogeneity.

eMethods 9: Complex-survey and reproducibility notes

County/district groups were derived from the first six digits of the public CHARLS identifier and community clusters from the first seven digits. The identifier meaning was checked against the 2011 CHARLS User Guide (pp. 22–23); every community in the eligible frame matched the official communityID file. These county groups are analysis pseudo-strata, not the original sampling strata. Earlier complete-case computations included 424 communities in 149 groups for transitions (275 degrees of freedom), with three singleton groups. Those computations and their average-contribution checks are historical sensitivities. Current core inference retains all 449 communities in 150 groups in the age-eligible frame (299 degrees of freedom; no singleton groups); communities outside an analysis domain contribute zero scores. Stage-specific sampling fractions were unavailable in the checked public files, so no finite-population correction was applied. See eMethods 13 for the scope and limitations of this approximation.

The final revision analysis used Python 3.12.7, pandas 3.0.5, NumPy 2.5.3, SciPy 1.18.1, statsmodels 0.15.0 and scikit-learn 1.9.1. Random seed 20260913 was set. The analysis log records SHA-256 hashes of the frozen inputs and derived pair file, all sample counts, weight diagnostics, bootstrap settings and verification gates. The final rerun reproduced the locked transition matrix and prognostic odds ratios from the preceding environment. The implementation preserves original predictor-missingness indicators across sequential data-construction steps. Compared with a weight model omitting those indicators, the largest absolute change in any pooled transition estimate was 0.04 percentage points. The 2018 feasibility audit found that harmonised r4 body mass index, grip strength and self-rated health fields required for the same phenotype definition were unavailable; 2018 was therefore not pooled with the 2011–15 analysis. Reconstruction from raw later-wave files was not undertaken, so this audit does not establish that later-wave harmonization is impossible.

A targeted reproducibility check on 24 September 2026 used the frozen pair-file hash recorded in the original analysis log and the matching Python environment. Refitting outcome-observation weights and the original prognostic models reproduced all 16 four-state transition estimates, six adjusted state odds ratios and four standardized ADL risks within numerical precision. This check did not reconstruct the phenotypes from original survey files or rerun the 200-replicate bootstrap and other sensitivity analyses.

Additional Reporting Tables

Supporting results are provided in eTables 5, 6.

eMethods 10: Revision joint-endpoint sensitivity analyses

These post hoc analyses were specified in the revision task packet before fitting on 25 September 2026. The 6,489 baseline ADL-free origin records contributed two approximately two-year intervals. Recorded death (next-wave interview status 5) took priority as a destination. Among known survivors (status 1 or 4), ADL=0 defined survival without disability and ADL>0 survival with disability. No recorded death had an observed subsequent ADL assessment. A destination remained unknown for 264 known survivors without ADL and 157 records with unresolved vital status; these 421 records were not coded as event-free. The 6,068 ascertained records comprised

4,944 alive without disability, 972 alive with disability and 152 deaths. Seventy-five lacked positive combined weights, including one death, leaving 5,993 records from 4,490 participants.

A logistic observation model was refitted to all 6,489 records for ascertainment of any destination, using the existing demographic, lifestyle, chronic-condition, CES-D, participation, joint-state, interval and original predictor-missingness terms. Predicted observation probabilities were bounded at 0.01–0.99. State-by-interval observation rates formed the stabilized numerator. Outcome-observation weights were truncated at their 1st/99th percentiles, multiplied by the existing phenotype-observation weight and interval-normalized sampling weight, and the positive combined weights were truncated again. The joint-outcome model converged; final weights ranged from 0.282 to 3.087, with effective sample size 4,494 (75.0% of 5,993). The weighted proportions in eTable 12 sum across mutually exclusive destinations. This analysis targets a measured-covariate-adjusted description of classified origin records; it does not identify outcomes for all records with missing origin phenotypes, and its observation assumptions are unverifiable.

The binary composite was death or disability versus alive without disability. Covariates matched the original prognostic models; survey Taylor covariance treated the fitted weights as fixed. Marginal predictions for each joint state were averaged over the weighted covariate distribution of the 5,993 analyzed records, with delta-method intervals. The estimand concerns status at the next wave: transient disability followed by recovery, or disability preceding death, cannot be separately ascertained. It is neither an event-history competing-risk model nor a causal contrast. Both constituent outcomes and the composite are reported to avoid obscuring their different clinical meanings.

For a separately defined survivor-conditional analysis, the observation model was refitted among 6,180 records known alive, of which 5,916 had subsequent ADL and 5,842 positive weights. The effective sample size was 4,424; the weight range was 0.287–3.030. The analysis excludes deaths rather than treating them as ordinary ADL nonresponse. It remains descriptive because conditioning on survival can itself induce selection.

The composite was additionally fitted with categorical education and unchanged weights. For the primary mortality, ADL and composite logistic models, an average nonsingleton-stratum contribution replaced zero contributions from singleton strata. The reimplemented zero-contribution covariance was checked against the original calculation before applying this alternative. All models converged. Direct S3-versus-S2 comparisons were subsequently computed as beta(S3) minus beta(S2), with variance Var(beta3)+Var(beta2)−2Cov(beta3,beta2). Four comparisons (mortality, ADL, composite, and chronic-count-adjusted ADL) formed one exploratory Holm family. Unadjusted 95% intervals use the survey-design t distribution. Each contrast and variance was verified against refitting with S2 as the reference state (eTable 25). The survivor model was subsequently promoted to the principal disability presentation in the 28 September revision for estimand clarity. Original-model results remain below for traceability; revised direct contrasts are in eTable 29. No bootstrap was rerun for the analyses in this section; subsequent nuisance-refit resampling is described in eMethods 12. No new external validation was undertaken, and the sensitivity intervals remain unadjusted; Holm-adjusted P values are provided specifically for the four direct discordant-state comparisons. Supporting results are provided in eTables 4, 12, 13, 14, 15, 16.

eMethods 11: Scientific-revision analyses and estimand prioritization

All analyses in this section were recorded in a revision task packet before fitting on 28 September 2026 and were post hoc. We retained the original study definitions and added the fixed alternatives listed below without selecting a specification by statistical significance. Already-computed known-survivor ADL results were promoted to the principal disability presentation; the original observation-weighted ADL model remains a legacy sensitivity. State associations, standardized probabilities and S3-versus-S2 contrasts were refitted and checked against prior estimates or a change of reference group. The revised four-comparison Holm family uses mortality, survivor ADL, composite and chronic-count-adjusted survivor ADL. It is distinct from the archived original-model family in eTable 25; neither correction covers every exploratory analysis.

For mortality, the eligible set was all observed-origin records; ascertainment was known vital status. For survivor ADL, eligibility additionally required no baseline ADL difficulty and known follow-up survival; ascertainment was observed follow-up ADL. For the joint endpoint, eligibility required no baseline ADL difficulty and ascertainment was recorded death or observed ADL among known survivors. Outcome-observation models included origin state, interval, demographic/lifestyle/health predictors and original missingness indicators. Constant indicators were omitted as redundant; predictions and weights reproduced the existing implementation. Probabilities were bounded to 0.01–0.99, with 1st/99th percentile truncation of observation weights and final combined weights. Counts of lower and upper bounding are reported separately: near-one probabilities indicate near-complete ascertainment rather than large inverse weights. Near-zero probabilities in sparse nonrespondent patterns still raise positivity concerns. No outcome or origin state was imputed.

Balance was evaluated within each state-by-interval cell, not between exposure groups. The target mean and standard deviation used the phenotype-observation weight times interval-normalized sampling weight among eligible records with positive base weights. The pre-observation-adjustment mean used those base weights among observed records; the adjusted mean used the final combined weights. Each standardized difference divided the observed-minus-target mean by the target standard deviation. Constant covariates were omitted. Because stabilized numerators vary by state and interval, pooled unstandardized differences are not a substitute for these within-cell diagnostics. eTable 27 reports cell maxima; the full covariate-level diagnostics are retained in the reproducibility outputs. Residual imbalance was not concealed or used to select a more favorable weighting specification.
Flexible-coding sensitivities added squared centered age and categorical education to both the outcome-observation and outcome models. The upstream phenotype-observation weights were unchanged, so this is not a full-pipeline specification check. Alternative deficit thresholds were 0.20 and 0.30 with capacity unchanged; an alternative capacity threshold was at least one impaired domain with deficit threshold 0.25. Observation weights were refitted after state reclassification. Complete phenotype person-waves remained 13,530; the state membership, transition-cell denominators and associations changed. Reclassification counts use all complete person-waves and are unweighted. Full 4-by-4 transition matrices are provided in eTable 34. Survivor-ADL outcome analyses included 5,842 records and 960 events under each definition; this does not mean the state-specific groups were identical.
Continuous-component models used the impaired-domain count (0–4) and deficit index divided by 0.10, first additively and then with their product, with the same survivor sample, weights and covariates. In the product model the component main effects are conditional on the other component being zero, not average effects. No corrected AUC, calibration or external validation was computed for these new models. All sensitivity intervals and the continuous-interaction P value are unadjusted and descriptive.
The transition singleton check retained the original ratio estimates and stratified-PSU score aggregation, including zero-score PSUs for domain estimates. Each singleton received the mean covariance-meat contribution across nonsingleton strata rather than zero. Original zero-contribution intervals were reproduced before comparison. Differences used a signed destination indicator within origin (S2: S1 minus S4; S3: S4 minus S1), allowing for the covariance of mutually exclusive destinations. Estimated weights remained fixed. These within-origin contrasts do not compare full distributions across origins. The origin-missingness profile uses unweighted available-case means and reports each variable's denominator; it is descriptive, not correction for missing origin states.
A further post hoc sparse-indicator sensitivity used a rule recorded before fitting: omit endpoint-observation missingness indicators with 1–9 flagged eligible records, while retaining every record, the underlying predictors, other covariates and upstream phenotype weights. This removed CES-D, smoking, alcohol and social-participation missingness indicators, each flagged in only one or two eligible records. In the survivor-ADL and composite samples, the sole record flagged for smoking missingness had an unobserved endpoint. The rule was not selected using outcome associations or statistical significance. Probability bounding at 0.01–0.99 and 1st/99th percentile weight truncation were unchanged. Outcome models were refitted with the resulting weights; Table 3 retains the original specification. eTable 35 reports all three endpoint models and all state comparisons. These checks address sensitivity to sparse dummy coding, not full-pipeline uncertainty or identification of missing origin phenotypes.
Input checks covered unique participant-wave/interval keys, score ranges, input hashes, model convergence and full-rank new observation-model matrices, reparameterized contrasts, and transition row sums. New models used Python 3.12.7, pandas 3.0.5, NumPy 2.5.3, SciPy 1.18.1 and statsmodels 0.15.0. No raw phenotype reconstruction, full-pipeline bootstrap, temporal or external validation was performed.

eMethods 12: Logical category identification and nuisance-refit resampling

Both analyses were added post hoc on 29 September 2026 with rules recorded before fitting. The original complete-domain analysis remains the reference specification; the expanded logical-classification analysis is not substituted because its results are more favorable. All previously observed joint states were required to remain unchanged.
For logical identification, k was the sum of observed impaired domains and m the number of unavailable domains among cognition, locomotion, vitality and sensory capacity. Since each missing domain could be zero or one, the total impaired-domain count lay in [k,k+m]. The original threshold of at least two was satisfied for every possible completion if $k \geq 2$; it was not satisfied for any completion if $k+m<2$. Other records remained unresolved. The deficit-index rule of at least 16 observed items and the 0.25 threshold were unchanged. Missing domains were never assigned values. Thus, this procedure identifies categories in some incomplete records but does not measure their full capacity profiles or identify records whose possible scores straddle the threshold. We used existing domain-level measurements rather than inferring additional domains from their partly observed constituent items.
Phenotype-observation and follow-up/outcome-observation weights were refitted after the expansion, retaining original covariates, probability bounds and weight truncation. The disability estimand remained conditional on known survival. All 16 transitions and all three endpoint models are reported. Current within-origin contrasts,

expanded transition cells and associations use the 449-community eligible-frame Taylor specification without singleton strata. Earlier variance conventions remain archived, not mixed into the updated eTables 37–38. Expanded cohorts have different denominators; changes in point estimates combine inclusion and refitted-weight effects and do not quantify a causal effect of missingness. The original main tables and flow chart describe the complete-domain analysis; eTable 36 separately records the expansion.

For resampling, the eligible 25,658 person-wave records were grouped by county/district analysis pseudo-stratum and village/community cluster using the checked identifier mapping. Within each stratum, we sampled its original number of PSUs with replacement, copying all eligible records of each selected PSU, including repeated participants and waves. The eligible resampling frame comprised 449 community clusters in 150 analysis pseudo-strata and contained no single-PSU strata. This differs from singleton counts in complete-case analytic subsets. Constant missingness indicators were omitted as redundant; all newly fitted model matrices were required to have full column rank, and all models were required to converge. Each replicate repeated wave-specific median predictor imputation (pooled fallback), phenotype-observation modeling, stabilized numerators, follow-up/outcome-observation modeling, survey-weight normalization, both truncation stages and outcome regression. Unresampled point estimates were checked against all 16 original transition cells and nine endpoint ORs before resampling. No replicate was excluded for an unfavorable estimate; failed fits, if any, are reported separately in the audit file. Seed: 20260929; attempted replicates: 200.

Percentile limits are the empirical 2.5th and 97.5th percentiles of successful estimates. This ordinary within-stratum n-PSU bootstrap is an exploratory refitting sensitivity, not the Rao–Wu rescaled bootstrap or a calibrated official survey replicate-weight procedure. Phenotype definitions, cognitive cutoffs, supplied survey weights and missing-domain patterns were held fixed. It does not propagate measurement-threshold estimation, original survey calibration, uncertainty in the latent missing values, and its finite number of replicates limits tail precision. It is not multiple imputation, a test of missing at random, predictive validation or external validation. It must not be used to claim all selection or full-pipeline uncertainty has been resolved.

eMethods 13: Eligible-frame variance integration and design sensitivity

The first six identifier digits distinguish county/district groups and the first seven identify communities (CHARLS 2011 User Guide, pp. 22–23; https://charls.pku.edu.cn/en/data/User2011.pdf). Official sampling documentation describes region/urban-rural/GDP sorting followed by systematic PPS county selection, then three communities per selected county. Its use of the term PSU for communities does not imply that counties were selected with certainty. County/district ID groups are therefore described here as analysis pseudo-strata, not recovered original sorting strata. Current core Taylor inference retains all 449 communities in the 25,658 age-eligible person-wave frame, in 150 groups, including zero-score communities outside the respective analysis. Cluster scores sum all included person-period contributions, preserving repeated-person clustering. There are 299 variance degrees of freedom and no singleton groups. Estimated weights and measurements remain fixed; standardized-risk delta intervals also condition on the empirical covariate distribution.

This post hoc variance integration changes neither sample inclusion, regression covariates, weights nor point estimates. Main Tables 2–4 and eTables 10–15, 28–30, 32, 34, 37–38 use updated inference. Earlier model/variance specifications remain explicitly archived in eTables 6, 16–25, 31 and 35. Tables containing counts, point estimates or diagnostics without intervals are unchanged. The 200-replicate nuisance-refit analysis in eTable 39 already used the 449-community frame; it was not rerun or relabelled as full original-design inference.

Five specifications were compared with fixed models and weights: the earlier analytic-subset county-pseudo-stratum/community calculation; the 449-community eligible frame; the official 450-community frame; unstratified community clustering; and unstratified county clustering. The latter two are sensitivity sandwiches, not guaranteed conservative bounds or recovered official design variances. Neither county pseudo-stratification nor unstratified county clustering reconstructs unknown joint inclusion probabilities, original systematic sampling or survey-weight calibration. The specifications were not selected according to significance. Focal contrasts and direct outcome comparisons are shown in eTables 40–41; state associations under alternative clustering are in eTable 42. Intervals remain unadjusted; Holm adjustment covers the four direct outcome contrasts separately within each classification/specification, not all exploratory comparisons across this supplement.

The score/Hessian implementation reproduced the previous covariance; county-cluster covariance was checked against an independent statsmodels sandwich implementation with the same $G/(G-1)$ correction. Multiplying all weights by seven left covariance unchanged. Current core and expanded-classification intervals reproduced a separately calculated design-sensitivity audit. Input hashes and all core point estimates were unchanged. These checks support implementation consistency, not absence of selection bias or independent external validation.

Supplemental Tables

eTable 1. Analytic flow

Stage	n	Detail
Eligible interviewed person-waves aged ≥60	25658	11619 participants
Joint phenotype observed	13530	52.7% of eligible records; 7785 participants
Records with an observed origin phenotype	8149	2011–13 and 2013–15 intervals
Subsequent phenotype observed	5002	Unweighted four-state transition denominator
Death before the next wave	230	Mutually exclusive from observed subsequent phenotype
Known alive without a complete subsequent phenotype	2706	Interview non-response or incomplete phenotype
Subsequent vital/phenotype status unresolved	211	Excluded from observed-transition analysis
Observed transitions with positive combined weight	4932	Effective sample size 3453
Observed phenotype or recorded death	5232	5002 observed phenotypes plus 230 deaths
Five-destination analysis, positive weights	5160	Includes 228 positive-weight deaths
Known subsequent vital status	7938	Includes records with an incomplete subsequent phenotype
Mortality logistic analysis, positive weights	7834	5702 participants; 228 deaths
Baseline ADL-free origin records	6489	Eligible for incident-ADL assessment
Observed subsequent ADL assessment	5916	Does not require an observed subsequent phenotype
ADL logistic analysis, positive weights	5842	4381 participants; 960 events
Baseline ADL-free records without follow-up ADL	573	Includes 152 recorded deaths; remaining 421 without an observed ADL assessment

Note. eFigure 1 presents the flow graphically.

eTable 2. Pooled analytic origin-record profile by joint state

Characteristic	Overall	S1	S2	S3	S4
Age, years, mean	67.59	66.56	69.29	66.98	69.60
Women, %	44.0	37.0	49.0	48.0	55.0
Chronic conditions, mean	1.74	1.04	1.04	3.36	3.05
Depressive-symptom score, mean	7.83	5.82	9.13	8.75	11.75
Any social participation, %	53.0	55.0	49.0	58.0	48.0
Origin records, n	8149	4219	1386	1117	1427

Characteristic	Overall	S1	S2	S3	S4
Unique participants, n	5924	3248	1278	978	1268

Note. Values other than counts use interval-normalised sampling weights. Counts are unweighted; participants can contribute to more than one state across intervals, so state-specific participant counts do not sum to the overall count.

eTable 3. Missingness by wave

Component	2011 missing, %	2013 missing, %	2015 missing, %
Cognition domain	33.73	34.12	28.12
Locomotion domain	0.99	1.55	1.09
Vitality domain	23.59	30.62	23.02
Sensory domain	7.94	10.44	8.07
Reduced-overlap deficit index	1.47	4.58	8.24
Four-state phenotype	44.98	51.45	45.34

Note. The phenotype required all four capacity domains plus a calculable deficit index. Cognition and vitality were the main contributors to phenotype incompleteness.

eTable 4. Source variables and analysis coding

Variable	Harmonized source	Analysis coding
Age	r*agey	Years; continuous linear term
Sex	ragender	1 male → 0; 2 female → 1
Education	raeduc_c	Codes 1–10; primary linear ordinal term, not schooling years
Partnered	r*mstat	1 married, 2 married/spouse absent, 3 partnered → 1; other valid codes → 0
Rural residence	h*rural	0 urban (reference); 1 rural
Ever smoking	r*smokev	0 no (reference); 1 yes
Past-year alcohol	r*drinkl	0 none (reference); 1 any alcohol in past year
Interval	Origin/next wave	2011–2013 reference; 2013–2015 indicator
ADL count	r*adlab_c	0–6 difficulties; baseline 0 required; follow-up >0 defines disability
Vital status	Next-wave r*iwstat	5 death; 1 or 4 known alive; other codes unresolved

Note. The asterisk denotes the wave number. The local file label is "Harmonized CHARLS Ver.D"; the corresponding documentation is Version D, June 2021. Education codes are: 1 no formal education/illiterate; 2 incomplete primary but literate; 3 Sishu; 4 elementary; 5 middle; 6 high school; 7 vocational; 8 associate/two- or three-year college; 9 bachelor; 10 postgraduate. Education is not inherently equally spaced; the composite analysis additionally used categorical education while retaining the same weights. Missing predictor values and original missingness flags were handled as described in eMethods 2. Source metadata were checked directly against H_CHARLS_D_Data.dta.

eTable 5. Original missingness of primary adjustment covariates

Variable	Origin records, n/N (%)	Mortality model, n/N (%)	ADL model, n/N (%)
Age	0/8149 (0.00)	0/7834 (0.00)	0/5842 (0.00)
Sex	0/8149 (0.00)	0/7834 (0.00)	0/5842 (0.00)
Education	0/8149 (0.00)	0/7834 (0.00)	0/5842 (0.00)
Partnered status	0/8149 (0.00)	0/7834 (0.00)	0/5842 (0.00)
Rural residence	0/8149 (0.00)	0/7834 (0.00)	0/5842 (0.00)
Ever smoking	1/8149 (0.01)	1/7834 (0.01)	0/5842 (0.00)
Alcohol use	2/8149 (0.02)	2/7834 (0.03)	2/5842 (0.03)

Note. Counts refer to missing values before median replacement, using the preserved original missingness indicators. N counts person-period records, not unique participants. These are covariate-missingness counts, not phenotype-missingness rates; the latter are reported in eTable 3.

eTable 6. Outcome-regression-unadjusted joint-state associations

Outcome	State versus S1	Weighted OR (95% CI)	N/events
Death before next wave	S2	1.41 (0.94–2.12)	7834/228
Death before next wave	S3	1.17 (0.73–1.90)	7834/228

Outcome	State versus S1	Weighted OR (95% CI)	N/events
Death before next wave	S4	3.06 (2.11–4.43)	7834/228
Incident ADL disability	S2	2.12 (1.68–2.67)	5842/960
Incident ADL disability	S3	2.31 (1.77–3.02)	5842/960
Incident ADL disability	S4	4.44 (3.49–5.66)	5842/960

Note. OR, odds ratio; CI, confidence interval. S1, higher capacity/lower deficits; S2, lower capacity/lower deficits; S3, higher capacity/higher deficits; S4, lower capacity/higher deficits. Models contained joint state alone and used the original outcome-specific combined sampling/selection weights and original analytic-subset Taylor variance method. Unadjusted refers only to omission of outcome-regression covariates; the weights still incorporate measured predictors of observation. These estimates supplement, and do not replace, the adjusted primary results in Table 3 and eTable 17.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 7. Weight diagnostics

Weight	n	Median	1st–99th percentile	Maximum	Effective n (%)
Phenotype IPW	8149	0.830	0.608–1.614	1.614	7700 (94.5)
Follow-up IPW	5002	0.967	0.727–1.576	1.577	4892 (97.8)
Combined selection and sampling weight	4932	0.667	0.249–3.498	3.500	3453 (70.0)

Note. Before weighting, the largest absolute standardised difference between observed and target records was 0.188. After weighting, all assessed absolute standardised differences were <0.10; the largest was 0.082 for age in the phenotype-observation model. Weighting cannot balance unmeasured determinants of missingness.

eTable 8. Worst-case bounds for focal transitions among records known alive

Origin	Destination	Known alive / destination missing, n	Weighted destination missing, %	Complete-destination estimate, %	Worst-case bound, %
S2	S1	1302 / 584	48.57	37.67	19.37–67.95
S2	S4	1302 / 584	48.57	17.73	9.12–57.69
S3	S1	1025 / 327	32.15	13.10	8.89–41.04
S3	S4	1025 / 327	32.15	30.95	21.00–53.15

Note. Base weights combined the origin phenotype-observation weight and interval-normalised sampling weight; follow-up observation weights were unavailable by definition for missing destinations. Lower bounds assign every missing destination away from the target state, and upper bounds assign every missing destination to it. These non-parametric bounds are partial-identification diagnostics rather than alternative adjusted estimates. They address missing destinations among records with observed origins and known survival, not the 47.3% of eligible person-waves lacking a complete phenotype at origin.

eTable 9. Reversal thresholds for the focal discordant-state orderings

Origin	Observed ordering	Weighted destination missing, %	Difference required among missing records to reverse the ordering
S2	P(S1) > P(S4)	48.57	P(S4) − P(S1) ≥21.12 percentage points
S3	P(S4) > P(S1)	32.15	P(S1) − P(S4) ≥37.66 percentage points

Note. Both thresholds are mathematically attainable. The focal ordering was retained after weighting for measured predictors; the attainable reversal thresholds nevertheless show that unmeasured selection could overturn it.

eTable 10. Selected discordant-state transitions by interval

Interval	Transition	Unweighted n/origin n	Weighted % (95% CI)
2011–13	S2→S1	142/330	42.86 (36.24–49.74)
2011–13	S2→S4	50/330	13.82 (9.83–19.08)
2011–13	S3→S1	48/284	17.27 (12.43–23.49)
2011–13	S3→S4	83/284	32.70 (25.87–40.34)
2013–15	S2→S1	136/397	32.89 (27.84–38.37)
2013–15	S2→S4	79/397	20.46 (16.20–25.50)

Interval	Transition	Unweighted n/origin n	Weighted % (95% CI)
2013–15	S3→S1	49/426	11.19 (7.75–15.90)
2013–15	S3→S4	126/426	31.17 (25.85–37.05)

Note. These interval estimates are descriptive. No multiplicity-adjusted formal test was specified for differences between intervals.
Current intervals use the 449-community eligible-frame covariance (299 degrees of freedom).

eTable 11. Weighted five-state destinations

Origin	Death, n; % (95% CI)	S1, %	S2, %	S3, %	S4, %
S1	80; 3.54 (2.70–4.63)	66.82	12.89	10.24	6.51
S2	43; 6.58 (4.65–9.25)	35.00	32.29	9.70	16.43
S3	32; 4.07 (2.61–6.30)	13.45	4.15	48.05	30.28
S4	75; 12.59 (9.78–16.07)	6.25	7.17	20.63	53.36

Note. Counts are unweighted. Percentages are row percentages after combined weighting and sum to 100 within origin state.
Current intervals use the 449-community eligible-frame covariance (299 degrees of freedom).

eTable 12. Three-destination death–disability distribution

State	Next-wave destination	Unweighted n/N	Weighted % (95% CI)
S1	Alive with disability	378/3614	11.05 (9.83–12.39)
S1	Alive without disability	3163/3614	86.71 (85.28–88.02)
S1	Death	73/3614	2.25 (1.71–2.95)
S2	Alive with disability	204/1035	20.73 (17.88–23.90)
S2	Alive without disability	798/1035	76.34 (73.08–79.32)
S2	Death	33/1035	2.93 (2.02–4.22)
S3	Alive with disability	161/741	22.24 (18.43–26.57)
S3	Alive without disability	561/741	75.33 (70.70–79.44)
S3	Death	19/741	2.44 (1.39–4.24)
S4	Alive with disability	217/603	34.39 (30.10–38.96)
S4	Alive without disability	360/603	60.51 (56.00–64.85)
S4	Death	26/603	5.09 (3.42–7.53)

Note. Counts and percentages both refer to the 5,993 records with observed destination and positive joint-outcome weights. Each state's three percentages sum to 100%, apart from rounding. Taylor intervals retain 449 community clusters in 150 analysis pseudo-strata (299 degrees of freedom), including zero-score communities. The denominator is baseline ADL-free records, not all origin records in the five-destination phenotype analysis. S1–S4 are defined in eMethods 1.

eTable 13. Adjusted death-or-disability composite

State	Adjusted OR (95% CI)	Standardized % (95% CI)	Difference vs S1, pp (95% CI)
S1	Reference	14.12 (12.64–15.59)	Reference
S2	1.60 (1.28–2.00)	20.63 (17.76–23.49)	6.51 (3.18–9.84)
S3	2.25 (1.74–2.92)	26.69 (22.16–31.22)	12.57 (7.93–17.21)
S4	3.53 (2.79–4.46)	36.01 (31.69–40.33)	21.89 (17.20–26.58)

Note. N=5,993 person-period records from 4,490 participants; 1,111 events (960 disability and 151 death). Comparator outcome: alive without disability. Models adjusted for age, sex, education, partnered status, rural residence, ever smoking, past-year alcohol use and interval. Standardized probabilities were averaged over the weighted analytic covariate distribution, with delta-method Taylor intervals. OR, odds ratio; CI, confidence interval; pp, percentage points. This interval-endpoint composite is not time-to-first-event cumulative incidence.
Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 14. ADL disability among known survivors

State	Adjusted OR (95% CI)	Standardized % (95% CI)
S1	Reference	11.94 (10.54–13.34)
S2	1.73 (1.36–2.20)	18.89 (16.05–21.72)
S3	2.43 (1.86–3.17)	24.52 (20.15–28.88)
S4	3.73 (2.90–4.81)	33.08 (28.69–37.47)

Note. Weights were refitted among 6,180 baseline ADL-free records known alive at the next wave. Of these, 5,916 had ADL data; 5,842 had positive weights, including 960 disability events. Deaths and unknown vital status were excluded before estimating ADL-observation weights.

Adjustment and standardization otherwise followed eTable 13. These are conditional descriptive associations, not causal effects among always-survivors.

Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 15. Categorical-education composite sensitivity

State vs S1	OR (95% CI)
S2	1.61 (1.29–2.01)
S3	2.28 (1.76–2.95)
S4	3.52 (2.78–4.46)

Note. Education codes entered as category indicators instead of a linear ordinal term in the outcome model. N=5,993, with 1,111 events. The same joint-outcome weights and other covariates were retained; this does not test education coding in the observation-weight models.

Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 16. Singleton-stratum variance sensitivity

Analysis / state vs S1	Zero-contribution CI	Average-contribution CI
Composite / S2	1.277–1.995	1.276–1.996
Composite / S3	1.743–2.917	1.741–2.920
Composite / S4	2.790–4.463	2.788–4.467
Original ADL / S2	1.344–2.186	1.343–2.188
Original ADL / S3	1.848–3.165	1.846–3.168
Original ADL / S4	2.865–4.783	2.862–4.787
Primary mortality / S2	0.631–1.540	0.630–1.542
Primary mortality / S3	0.763–2.062	0.762–2.066
Primary mortality / S4	1.524–3.524	1.522–3.529

Note. Point estimates are unchanged. Each listed model had one singleton stratum. The alternative adds the mean covariance-meat contribution of nonsingleton strata for each singleton before applying the model bread matrix, retaining the original design degrees of freedom. All nine comparisons retained the same inclusion/exclusion of OR=1. Weights and imputed values remain fixed; these checks alone do not validate the complete survey/selection uncertainty; transition-cell extensions are in eTable 31.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 17. Adjusted joint-state associations

Outcome	State versus S1	OR (95% CI)	P value	N/events
Death before next wave	S2	0.99 (0.63–1.54)	0.949	7834/228
Death before next wave	S3	1.25 (0.76–2.06)	0.370	7834/228
Death before next wave	S4	2.32 (1.52–3.52)	<0.001	7834/228
Incident ADL disability	S2	1.71 (1.34–2.19)	<0.001	5842/960
Incident ADL disability	S3	2.42 (1.85–3.16)	<0.001	5842/960
Incident ADL disability	S4	3.70 (2.86–4.78)	<0.001	5842/960

Note. Models adjusted for age, sex, education, partnered status, rural residence, ever smoking, past-year alcohol use and interval. Median-replaced covariates were used in the original models, and confidence intervals used stratified-primary-sampling-unit Taylor linearisation. Of 230 recorded deaths, two lacked a positive combined weight, leaving 228 events. For ADL, 6,489 baseline ADL-free origin records were eligible, 5,916 had observed subsequent ADL, and 5,842 records from 4,381 participants had positive outcome-specific weights; the analysis did not condition on subsequent joint-phenotype completeness. Death was not incorporated as a competing outcome in this original ADL model; its observation model included deaths as absent ADL assessments. The resulting estimand is descriptive within the weighted ADL-observed sample, not baseline-population disability cumulative incidence or a causal survivor effect. Alternative joint-endpoint and known-survivor analyses are reported in eMethods 10.

The current survivor-conditional standardized risks in main Table 4, rather than the legacy ADL model in this table, were obtained by setting every record in turn to S1, S2, S3 and S4 while retaining its analysis covariates (including single-imputed values), predicting from the primary weighted logistic model and averaging with the outcome-specific analysis weights. Risk differences used S1 as the reference. Current delta-method intervals use the eligible-frame covariance, conditional on estimated weights and the empirical covariate distribution.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 18. Participant-cluster bootstrap internal validation

Outcome	Model	Apparent AUC	Corrected AUC	Apparent Brier	Corrected Brier
Death	Covariates only	0.727	0.713	0.03131	0.03175
Death	+ capacity proxy	0.736	0.720	0.03125	0.03171
Death	+ deficit burden	0.745	0.730	0.03113	0.03158
Death	+ capacity and deficit burden	0.745	0.729	0.03110	0.03157
Death	+ four joint states	0.747	0.729	0.03105	0.03154
Incident ADL disability	Covariates only	0.611	0.603	0.13970	0.14071
Incident ADL disability	+ capacity proxy	0.646	0.638	0.13793	0.13904
Incident ADL disability	+ deficit burden	0.663	0.658	0.13592	0.13699
Incident ADL disability	+ capacity and deficit burden	0.678	0.671	0.13503	0.13621
Incident ADL disability	+ four joint states	0.678	0.671	0.13501	0.13628

Note. Internal validation used 200 bootstrap samples drawn with replacement at participant level (5,702 participants for mortality and 4,381 for incident ADL disability). All person-period records belonging to a selected participant remained in the resampled cluster. Each model was refitted in every bootstrap sample; weighted AUC and Brier scores were calculated in that sample and in the original estimation sample. Corrected AUC equals apparent AUC minus mean optimism; corrected Brier equals apparent Brier plus mean optimism. All 2,000 bootstrap model fits converged. The imputed covariates and previously estimated weights were fixed across resamples. Participants were sampled without reproducing the county-pseudo-stratum/community analysis specification, and the observation-weight models were not refitted. Thus, the correction estimates conditional in-sample optimism rather than full-pipeline survey-design uncertainty, and is not temporal or external validation. The additive-component and four-state models remained almost identical, showing no appreciable numerical gain from the interaction parameter; this is not a formal equivalence or interaction test.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 19. Complete-covariate sensitivity of fully adjusted joint-state associations

Outcome	State versus S1	OR (95% CI)	P value	N/events
Death before next wave	S2	0.98 (0.63–1.52)	0.921	7831/228
Death before next wave	S3	0.97 (0.54–1.75)	0.919	7831/228
Death before next wave	S4	1.84 (1.15–2.96)	0.011	7831/228
Incident ADL disability	S2	1.70 (1.33–2.17)	<0.001	5840/960
Incident ADL disability	S3	1.72 (1.22–2.44)	0.002	5840/960
Incident ADL disability	S4	2.73 (1.98–3.76)	<0.001	5840/960

Note. This analysis excluded records with any originally missing value for age, sex, education, partnered status, rural residence, ever smoking, past-year alcohol use or chronic-condition count and included chronic-condition count. It was consistent with the corresponding full-adjustment estimates in eTable 20.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 20. Sensitivity to additional adjustment for chronic-condition count

Outcome	State versus S1	Primary OR (95% CI)	Chronic-count-adjusted OR (95% CI)
Death before next wave	S2	0.99 (0.63–1.54)	0.98 (0.63–1.52)
Death before next wave	S3	1.25 (0.76–2.06)	0.97 (0.54–1.75)
Death before next wave	S4	2.32 (1.52–3.52)	1.84 (1.15–2.96)
Incident ADL disability	S2	1.71 (1.34–2.19)	1.69 (1.33–2.16)
Incident ADL disability	S3	2.42 (1.85–3.16)	1.72 (1.22–2.44)
Incident ADL disability	S4	3.70 (2.86–4.78)	2.73 (1.98–3.76)

Note. The primary model omitted chronic-condition count because all 12 conditions were direct components of the deficit index. The sensitivity model added the count to estimate associations beyond measured chronic disease burden; attenuation was expected, but the ordering of ADL associations remained unchanged.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 21. Sensitivity to ADL observation-weight truncation

Truncation	S2 versus S1, OR (95% CI)	S3 versus S1, OR (95% CI)	S4 versus S1, OR (95% CI)
1st/99th percentiles (primary)	1.71 (1.34–2.19)	2.42 (1.85–3.16)	3.70 (2.86–4.78)
2.5th/97.5th percentiles	1.71 (1.35–2.16)	2.39 (1.84–3.10)	3.69 (2.87–4.73)
Untruncated	1.77 (1.37–2.28)	2.35 (1.66–3.33)	4.17 (3.06–5.69)

Note. All models used 5,842 positive-weight records with 960 incident-ADL events and the primary covariate set. The direction and ordering were unchanged.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 22. Core weighted transition estimates under alternative definitions

Specification	S2→S1, % (95% CI)	S2→S4, % (95% CI)	S3→S1, % (95% CI)	S3→S4, % (95% CI)
Wave-specific cognitive threshold	37.21 (32.86–41.76)	17.58 (14.64–20.98)	13.89 (10.76–17.75)	31.85 (27.73–36.27)
Fixed 2011 cognitive threshold (7.07)	36.62 (32.34–41.12)	17.70 (14.74–21.11)	13.89 (10.76–17.75)	32.84 (28.61–37.38)
All 19 deficit items observed	39.62 (34.53–44.94)	16.72 (13.47–20.57)	15.14 (11.58–19.54)	29.88 (25.39–34.80)

Note. The fixed-threshold results reduce concern that wave-specific cognitive standardisation alone generated the principal transition pattern. Requiring complete deficit items reduced the discordant-origin sample sizes but preserved the ordering of transitions.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 23. Five-domain sensitivity using depressive symptoms as a psychological proxy

Specification	Complete phenotype, n/N (%)	S2→S1, % (95% CI)	S2→S4, % (95% CI)	S3→S1, % (95% CI)	S3→S4, % (95% CI)
Four-domain primary definition	13530/25658 (52.73)	37.21 (32.86–41.76)	17.58 (14.64–20.98)	13.89 (10.76–17.75)	31.85 (27.73–36.27)
Five-domain sensitivity definition	13527/25658 (52.72)	34.02 (30.28–37.96)	19.72 (16.77–23.05)	13.20 (9.74–17.64)	38.13 (33.62–42.86)

Note. Psychological impairment was CES-D-10 ≥12, a threshold validated in older Chinese people (3). Lower capacity remained at least two impaired domains, now among cognition, locomotion, vitality, sensory and psychological proxies. The focal ordering was preserved. Across all 16 transition cells, the maximum absolute difference from the primary definition was 6.28 percentage points (S3→S4). The phenotype-observation model excluded the CES-D missingness indicator because CES-D observation was mechanically required by this alternative outcome definition; other non-constant predictor-missingness indicators were retained. CES-D-10 captures depressive symptoms and does not constitute a complete WHO psychological-capacity assessment; this analysis therefore reduces, but does not remove, the measurement limitation.

Additional analyses defined lower capacity as at least one impaired domain, broadened sensory impairment to fair/poor, changed the deficit threshold to 0.20 and used a conventional 34-item frailty index at thresholds of 0.25 and 0.20. These definitions changed state prevalence and model denominators, but the S2-to-S4 exploratory interaction remained positive in most specifications. Because interaction estimates varied across estimands and functional forms, these analyses do not convert it into a confirmatory finding.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 24. Exploratory interaction models

Origin and target	Term	OR (95% CI)	P value	N/events
S2→S4	CES-D per five points	0.95 (0.72–1.25)	0.699	718/127
S2→S4	Any social participation	0.42 (0.16–1.13)	0.085	718/127
S2→S4	CES-D × participation	1.66 (1.07–2.58)	0.025	718/127
S3→S1	CES-D per five points	0.78 (0.54–1.13)	0.184	698/96
S3→S1	Any social participation	1.45 (0.47–4.51)	0.521	698/96
S3→S1	CES-D × participation	0.94 (0.56–1.58)	0.810	698/96

Note. At CES-D scores of 0, 10 and 20, linear-model risk differences for participation versus non-participation in S2-to-S4 were −8.99 (−19.53 to 1.56), 1.97 (−3.93 to 7.87) and 18.40 (1.26–35.54) percentage points. Exact high scores were sparse, so these point contrasts should not be read as a dose-response curve.

The quadratic model gave a joint two-degree-of-freedom interaction P=0.034. Its participation risk differences were −22.13 (−39.43 to −4.84), 8.22 (−0.20 to 16.65) and 11.00 (−6.90 to 28.90) percentage points at CES-D 0, 10 and 20. eFigure 2 shows the standardised curves with confidence bands and the observed CES-D density. The curve changes direction and becomes imprecise at the tails.

An unweighted multinomial model used persistence as the reference. For S2-to-S4 versus S2 persistence, the interaction relative risk ratio was 1.38 (0.92–2.07; P=0.121). For S3-to-S1 versus S3 persistence, it was 1.00 (0.61–1.62; P=0.988). The local multinomial estimator did not support the combined survey/selection weights and Taylor covariance; this analysis therefore addresses a different, unweighted estimand and is retained only as a sensitivity check.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 25. Original-model direct S3-versus-S2 outcome contrasts

Model	OR (95% CI)	Raw P	Holm P	Events/N
Primary mortality	1.27 (0.70–2.33)	0.432	0.864	228/7834
Original ADL	1.41 (1.03–1.92)	0.030	0.093	960/5842
Death or disability	1.41 (1.05–1.90)	0.023	0.093	1111/5993
ADL additionally adjusted for chronic-condition count	1.02 (0.70–1.47)	0.930	0.930	960/5842

Note. S3 is higher capacity/higher deficit burden and S2 lower capacity/lower deficit burden. Positive ORs above one favor higher outcome odds in S3. All four comparisons are post hoc; Holm adjustment covers these four tests only, not the complete manuscript. Confidence intervals are unadjusted. Original models, covariates and outcome-specific weights were retained. Chronic-condition count is a component of the deficit exposure, so its addition changes the comparison and is not a definitive confounding correction. N denotes person-period records.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 26. Endpoint-specific weight and positivity diagnostics

Model	Eligible / observed / analysis n	ESS (%)	P<.01 / P>.99, n	IPW / combined truncated n	Final weight range
Mortality	8149 / 7938 / 7834	5851 (74.7)	0 / 3652	80 / 158	0.283–3.221

Model	Eligible / observed / analysis n	ESS (%)	P<.01 / P>.99, n	IPW / combined truncated n	Final weight range
Survivor ADL	6180 / 5916 / 5842	4424 (75.7)	1 / 126	119 / 118	0.287–3.030
Death or disability	6489 / 6068 / 5993	4494 (75.0)	1 / 4	122 / 120	0.282–3.087

Note. Probability counts refer to unbounded predictions. No observed record had a prediction below .01. Bounding and truncation are separate operations. ESS is a weight-only effective sample size, not a count of independent people. All models converged; sparse predictions do not establish adequate positivity.

eTable 27. State-by-interval ascertainment and balance

Model	State	Interval	Observed / eligible	Observed %	Max absolute SMD before / after
Mortality	S1	2011-2013	2076/2120	97.9	0.016 / 0.031
Mortality	S1	2013-2015	2044/2099	97.4	0.028 / 0.015
Mortality	S2	2011-2013	725/740	98.0	0.016 / 0.017
Mortality	S2	2013-2015	634/646	98.1	0.017 / 0.019
Mortality	S3	2011-2013	466/484	96.3	0.025 / 0.054
Mortality	S3	2013-2015	611/633	96.5	0.035 / 0.076
Mortality	S4	2011-2013	642/667	96.3	0.032 / 0.051
Mortality	S4	2013-2015	740/760	97.4	0.046 / 0.064
Survivor ADL	S1	2011-2013	1789/1898	94.3	0.022 / 0.038
Survivor ADL	S1	2013-2015	1798/1876	95.8	0.022 / 0.022
Survivor ADL	S2	2011-2013	526/544	96.7	0.015 / 0.022
Survivor ADL	S2	2013-2015	486/491	99.0	0.007 / 0.016
Survivor ADL	S3	2011-2013	309/323	95.7	0.054 / 0.116
Survivor ADL	S3	2013-2015	425/433	98.2	0.029 / 0.102
Survivor ADL	S4	2011-2013	254/272	93.4	0.034 / 0.110
Survivor ADL	S4	2013-2015	329/343	95.9	0.034 / 0.106
Death or disability	S1	2011-2013	1822/1971	92.4	0.038 / 0.043
Death or disability	S1	2013-2015	1839/1965	93.6	0.044 / 0.022
Death or disability	S2	2011-2013	547/577	94.8	0.027 / 0.019
Death or disability	S2	2013-2015	498/511	97.5	0.019 / 0.020
Death or disability	S3	2011-2013	317/342	92.7	0.040 / 0.108
Death or disability	S3	2013-2015	436/458	95.2	0.052 / 0.101
Death or disability	S4	2011-2013	267/298	89.6	0.081 / 0.114
Death or disability	S4	2013-2015	342/367	93.2	0.095 / 0.138

Note. Balance includes listed observation predictors and varying missingness indicators; target and scaling are defined in eMethods 11. This is observation-selection balance, not balance between S1–S4. A maximum below .10 does not prove absence of selection bias.

eTable 28. Definition sensitivity of focal transitions and disability associations

Definition	Reclassified person-waves	S2→S1 / S2→S4 / S3→S1 / S3→S4, %	Survivor ADL ORs S2 / S3 / S4 vs S1
Original	0 (0.00%)	37.21 / 17.58 / 13.89 / 31.85	1.73 / 2.43 / 3.73
DI cutoff .20	1904 (14.07%)	34.38 / 23.93 / 11.25 / 30.87	1.75 / 2.47 / 3.75

Definition	Reclassified person-waves	S2→S1 / S2→S4 / S3→S1 / S3→S4, %	Survivor ADL ORs S2 / S3 / S4 vs S1
DI cutoff .30	1576 (11.65%)	38.78 / 16.50 / 14.41 / 32.93	1.85 / 3.12 / 4.04
Capacity >=1	4698 (34.72%)	20.49 / 20.04 / 5.45 / 52.86	1.51 / 1.86 / 3.74

Note. All definitions have 13,530 complete person-waves and 5,842 positive-weight survivor-ADL records with 960 events. State-specific groups change. Reclassification denominator is 13,530. Full transition intervals appear in eTable 34; disability intervals are reported below.
DI cutoff .20: S2 1.75 (1.34–2.29). DI cutoff .20: S3 2.47 (1.96–3.13). DI cutoff .20: S4 3.75 (2.92–4.82). DI cutoff .30: S2 1.85 (1.48–2.30). DI cutoff .30: S3 3.12 (2.30–4.23). DI cutoff .30: S4 4.04 (3.01–5.43). Capacity >=1: S2 1.51 (1.17–1.95). Capacity >=1: S3 1.86 (1.19–2.91). Capacity >=1: S4 3.74 (2.80–4.98).
Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 29. Revised direct discordant-state outcome contrasts

Model	S3 vs S2 OR (95% CI)	Raw P	Holm P
Mortality	1.27 (0.70–2.33)	0.432	0.864
Survivor ADL	1.40 (1.03–1.91)	0.032	0.095
Death or disability	1.41 (1.05–1.90)	0.024	0.094
Survivor ADL + chronic count	1.00 (0.69–1.45)	0.992	0.992

Note. Four post hoc comparisons form one Holm family; intervals remain unadjusted. The original-model family is preserved in eTable 25. This family uses survivor-conditional rather than original ADL weights.
Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 30. Continuous-component survivor-ADL models

Model	Term	OR (95% CI)	Unadjusted P
Additive	ic_domains_impaired	1.27 (1.15–1.40)	<.001
Additive	di10	1.54 (1.42–1.67)	<.001
Interaction	ic_domains_impaired	1.21 (1.00–1.46)	0.051
Interaction	di10	1.49 (1.31–1.71)	<.001
Interaction	ic_domains_impaired:di10	1.02 (0.95–1.10)	0.533

Note. di10 is deficit index per .10; impaired-domain count is per one domain. Product-model main effects apply at zero of the other component. Same covariates as Table 3; N=5,842, 960 events. These are exploratory associative models, not validated prediction tools.
Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 31. Transition singleton correction and within-origin probability contrasts

Cell or contrast	Estimate, % or pp	Zero-contribution 95% CI	Average-contribution 95% CI
S1→S1	69.11	66.91–71.22	66.89–71.24
S1→S2	13.53	11.98–15.24	11.97–15.26
S1→S3	10.58	9.29–12.02	9.28–12.04
S1→S4	6.79	5.76–7.98	5.75–8.00
S2→S1	37.21	32.86–41.76	32.82–41.81
S2→S2	34.89	30.72–39.30	30.68–39.34
S2→S3	10.32	7.89–13.40	7.87–13.43
S2→S4	17.58	14.64–20.98	14.61–21.02
S3→S1	13.89	10.76–17.75	10.74–17.79
S3→S2	4.28	2.60–6.98	2.59–7.02
S3→S3	49.97	45.53–54.41	45.49–54.45
S3→S4	31.85	27.73–36.27	27.69–36.32
S4→S1	6.94	5.24–9.14	5.22–9.17
S4→S2	8.34	6.23–11.08	6.21–11.11
S4→S3	23.25	20.24–26.57	20.21–26.60
S4→S4	61.47	57.61–65.19	57.57–65.23
S2: S1 minus S4	19.62	13.22–26.02	13.16–26.08

Cell or contrast	Estimate, % or pp	Zero-contribution 95% CI	Average-contribution 95% CI
S3: S4 minus S1	17.96	11.62–24.30	11.56–24.36

Note. Percentages for cells and percentage points for differences. All use 424 PSUs, 149 strata, 275 design degrees of freedom and three singleton strata. Point estimates are unchanged. Intervals are descriptive and fixed-weight; no full-distribution or causal transition test is implied.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 32. Flexible age and education in observation and outcome models

Outcome	State vs S1	OR (95% CI)
Mortality	S2	1.00 (0.64–1.54)
Mortality	S3	1.22 (0.72–2.07)
Mortality	S4	2.25 (1.49–3.41)
Survivor ADL	S2	1.74 (1.37–2.22)
Survivor ADL	S3	2.45 (1.88–3.21)
Survivor ADL	S4	3.71 (2.87–4.78)
Death or disability	S2	1.61 (1.29–2.01)
Death or disability	S3	2.28 (1.76–2.96)
Death or disability	S4	3.48 (2.75–4.41)

Note. Squared centered age and categorical education were used in both endpoint-observation and outcome models. Upstream phenotype weights were retained. Mortality N=7,834; survivor ADL N=5,842; composite N=5,993. This is not a full-pipeline functional-form check.

Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 33. Measured profile by origin-phenotype availability

Variable	Origin observed?	Records	Variable observed n	Unweighted mean
age	True	8149	8149	67.16
age	False	7665	7665	69.67
female	True	8149	8149	0.43
female	False	7665	7665	0.57
education	True	8149	8149	3.20
education	False	7665	7661	2.47
multimorbidity	True	8149	8149	1.69
multimorbidity	False	7665	7253	1.68
cesd10	True	8149	8147	8.02
cesd10	False	7665	6090	9.25
social	True	8149	8148	0.52
social	False	7665	6310	0.42

Note. Eligible 2011/2013 person-waves only, not all three study waves. Means use available values separately for each variable; binary means are proportions. The profile does not identify unobserved origin states.

eTable 34. Full transition matrices under alternative definitions

Definition	Origin	To S1, % (95% CI)	To S2, % (95% CI)	To S3, % (95% CI)	To S4, % (95% CI)
Original	S1	69.11 (66.90–71.23)	13.53 (11.98–15.24)	10.58 (9.26–12.05)	6.79 (5.75–7.99)
Original	S2	37.21 (32.86–41.77)	34.89 (30.72–39.30)	10.32 (7.89–13.40)	17.58 (14.63–20.98)
Original	S3	13.89 (10.77–17.75)	4.28 (2.60–6.99)	49.97 (45.53–54.41)	31.85 (27.69–36.32)
Original	S4	6.94 (5.21–9.19)	8.34 (6.16–11.20)	23.25 (20.13–26.70)	61.47 (57.52–65.27)
DI cutoff .20	S1	65.36 (62.96–67.69)	10.31 (8.86–11.97)	16.16 (14.44–18.05)	8.16 (6.93–9.60)
DI cutoff .20	S2	34.38 (29.32–39.81)	25.39 (21.25–30.01)	16.30 (12.81–20.52)	23.93 (20.05–28.30)
DI cutoff .20	S3	11.25 (9.18–13.70)	2.61 (1.57–4.29)	55.28 (51.86–58.65)	30.87 (27.66–34.27)

Definition	Origin	To S1, % (95% CI)	To S2, % (95% CI)	To S3, % (95% CI)	To S4, % (95% CI)
DI cutoff .20	S4	6.05 (4.49–8.12)	5.09 (3.69–6.97)	26.63 (23.81–29.66)	62.23 (58.89–65.45)
DI cutoff .30	S1	71.37 (69.22–73.43)	16.36 (14.69–18.19)	7.18 (6.12–8.40)	5.08 (4.27–6.05)
DI cutoff .30	S2	38.78 (35.10–42.59)	39.11 (35.28–43.08)	5.61 (4.17–7.53)	16.50 (14.01–19.33)
DI cutoff .30	S3	14.41 (10.82–18.93)	6.53 (3.88–10.81)	46.13 (39.86–52.52)	32.93 (27.19–39.23)
DI cutoff .30	S4	8.31 (6.12–11.19)	11.64 (8.84–15.18)	20.00 (16.34–24.24)	60.05 (55.10–64.80)
Capacity >=1	S1	48.08 (44.85–51.34)	36.53 (33.56–39.60)	5.80 (4.49–7.47)	9.59 (7.79–11.74)
Capacity >=1	S2	20.49 (18.46–22.68)	56.94 (54.06–59.77)	2.53 (1.78–3.60)	20.04 (18.03–22.21)
Capacity >=1	S3	5.45 (3.33–8.79)	10.31 (6.27–16.48)	31.38 (24.28–39.48)	52.86 (45.10–60.49)
Capacity >=1	S4	4.33 (2.99–6.24)	12.39 (10.23–14.93)	5.46 (4.30–6.90)	77.82 (74.77–80.59)

Note. Each row sums to 100% apart from rounding. Selection weights were refitted for each alternative state definition. All use 4,932 positive-weight observed-transition records; cell membership changes. Intervals use the current eligible-frame convention; archived eTable 31 evaluates earlier analytic-subset singleton corrections.

Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 35. Sparse missingness-indicator sensitivity

Endpoint	Eligible / analysis / events	Minimum P(observed)	P<.01 / P>.99, n	S2 vs S1 OR (95% CI)	S3 vs S1 OR (95% CI)	S4 vs S1 OR (95% CI)
Mortality	8149 / 7834 / 228	0.662	0 / 3651	0.99 (0.63–1.54)	1.25 (0.76–2.06)	2.32 (1.52–3.52)
Survivor ADL	6180 / 5842 / 960	0.816	0 / 118	1.73 (1.36–2.20)	2.43 (1.86–3.17)	3.73 (2.90–4.81)
Death or disability	6489 / 5993 / 1111	0.699	0 / 0	1.60 (1.28–1.99)	2.25 (1.74–2.92)	3.53 (2.79–4.46)

Note. Post hoc, fixed-rule omission of missingness indicators flagged in 1–9 eligible records from endpoint-observation models only; no records or underlying predictors were removed. Probability counts use unbounded predictions in eligible samples. The maximum absolute state-OR change relative to Table 3 was 0.000575. Weight-only effective sample sizes were 5,851, 4,425 and 4,496, respectively. Covariates, variance conventions and upstream phenotype weights were retained. Intervals are unadjusted and descriptive. No full-pipeline bootstrap or external validation was performed.

Archived specification. This table retains its previously reported model and variance convention for traceability; it is not a source of the current eligible-frame intervals. Current core estimates are in main Tables 2–4 and eMethods 13.

eTable 36. Logical identification of categories in incomplete records

Wave	Eligible	Complete-domain category	Additional logically identified	Total classified	Unclassified
2011	7290	4011	1706	5717	1573
2013	8524	4138	2055	6193	2331
2015	9844	5381	2026	7407	2437
All	25658	13530	5787	19317	6341

Note. These are person-wave counts, not independent people. The original missing-measurement rate remains 47.3%; the expanded unclassified-category rate is 24.7%. No missing domains were imputed.

eTable 37. Transitions after including logically identified categories

Origin	To S1, % (95% CI)	To S2, % (95% CI)	To S3, % (95% CI)	To S4, % (95% CI)
S1	70.48 (68.74–72.17)	12.47 (11.32–13.71)	9.94 (8.83–11.16)	7.11 (6.29–8.03)
S2	25.46 (23.04–28.04)	40.95 (38.22–43.73)	6.11 (4.92–7.57)	27.48 (24.78–30.36)
S3	12.98 (10.52–15.92)	4.22 (2.87–6.17)	48.79 (45.03–52.56)	34.01 (30.50–37.70)
S4	4.13 (3.24–5.26)	10.16 (8.49–12.12)	13.52 (11.83–15.41)	72.19 (69.70–74.55)

Note. N=8,604 positive-weight transitions from 11,910 eligible origin records. All original observed states were unchanged. Eligible-frame covariance with fixed estimated weights. S2 (S1 minus S4) difference −2.02 pp (95% CI −6.55 to 2.51); S3 (S4 minus S1) 21.03 pp (15.86 to 26.19). These differences are within origin, not a comparison of whole transition distributions.

Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 38. Outcome associations after logical identification

Endpoint	Records / events	S2 vs S1 OR (95% CI)	S3 vs S1 OR (95% CI)	S4 vs S1 OR (95% CI)
Mortality	11393 / 469	1.09 (0.76–1.56)	1.14 (0.72–1.82)	3.31 (2.42–4.53)
Survivor ADL	7960 / 1461	2.13 (1.74–2.61)	2.52 (1.98–3.22)	4.94 (4.07–6.00)
Death or disability	8212 / 1713	1.92 (1.58–2.34)	2.28 (1.80–2.89)	4.62 (3.85–5.54)

Note. Same outcome covariates as Table 3, with refitted phenotype and endpoint weights in expanded samples. Fixed-weight Taylor intervals are unadjusted and descriptive. N denotes person-period records; events include repeated intervals from individuals. No comparison establishes that selection bias was removed.

Updated inference. Fixed-weight Taylor intervals use 449 communities, 150 analysis pseudo-strata and 299 degrees of freedom; no singleton correction is needed.

eTable 39. Original-specification nuisance-refit bootstrap sensitivity

Estimate	Original estimate	Bootstrap 2.5th percentile	Bootstrap 97.5th percentile	Successful replicates
S1->S1	69.11	67.30	70.85	200
S1->S2	13.53	12.31	15.05	200
S1->S3	10.58	9.43	11.45	200
S1->S4	6.79	5.83	7.67	200
S2->S1	37.21	33.58	41.05	200
S2->S2	34.89	31.85	38.12	200
S2->S3	10.32	8.10	12.73	200
S2->S4	17.58	15.11	20.09	200
S3->S1	13.89	11.21	16.64	200
S3->S2	4.28	2.68	5.97	200
S3->S3	49.97	46.00	53.48	200
S3->S4	31.85	28.28	35.45	200
S4->S1	6.94	5.30	8.49	200
S4->S2	8.34	6.53	10.49	200
S4->S3	23.25	20.74	25.80	200
S4->S4	61.47	58.41	64.57	200
S2:S1-S4	19.62	13.74	24.79	200
S3:S4-S1	17.96	12.95	22.57	200
Mortality:S2	0.99	0.69	1.43	200
Mortality:S3	1.25	0.75	1.95	200
Mortality:S4	2.32	1.62	3.36	200
Survivor ADL:S2	1.73	1.38	2.09	200
Survivor ADL:S3	2.43	1.89	3.04	200
Survivor ADL:S4	3.73	3.05	4.37	200
Death or disability:S2	1.60	1.30	1.91	200
Death or disability:S3	2.25	1.76	2.70	200
Death or disability:S4	3.53	2.94	4.18	200

Note. Seed 20260929; 200/200 successful replicates; 0 failed. 449 community clusters in 150 analysis pseudo-strata, with no singleton groups. Transition cells are percentages; within-origin contrasts percentage points; outcome estimates ORs. Resampling refits medians and weights but fixes phenotypes and supplied survey weights. This is not full-design uncertainty or predictive validation; see eMethods 12.

eTable 40. Focal transition contrasts under alternative variance specifications

Classification	Within-origin contrast	Variance specification	Difference, pp (95% CI)
Complete-domain	S2: S1 minus S4	Earlier subset	19.62 (13.22–26.02)
Complete-domain	S2: S1 minus S4	Eligible 449	19.62 (13.21–26.03)
Complete-domain	S2: S1 minus S4	Official 450	19.62 (13.21–26.03)
Complete-domain	S2: S1 minus S4	Community clusters	19.62 (13.23–26.02)
Complete-domain	S2: S1 minus S4	County clusters	19.62 (13.21–26.03)
Complete-domain	S3: S4 minus S1	Earlier subset	17.96 (11.62–24.30)
Complete-domain	S3: S4 minus S1	Eligible 449	17.96 (11.56–24.36)
Complete-domain	S3: S4 minus S1	Official 450	17.96 (11.56–24.36)
Complete-domain	S3: S4 minus S1	Community clusters	17.96 (11.83–24.09)
Complete-domain	S3: S4 minus S1	County clusters	17.96 (12.36–23.55)
Logical expansion	S2: S1 minus S4	Earlier subset	-2.02 (-6.54–2.50)
Logical expansion	S2: S1 minus S4	Eligible 449	-2.02 (-6.55–2.51)
Logical expansion	S2: S1 minus S4	Official 450	-2.02 (-6.55–2.51)
Logical expansion	S2: S1 minus S4	Community clusters	-2.02 (-6.51–2.47)
Logical expansion	S2: S1 minus S4	County clusters	-2.02 (-6.48–2.44)
Logical expansion	S3: S4 minus S1	Earlier subset	21.03 (15.91–26.15)
Logical expansion	S3: S4 minus S1	Eligible 449	21.03 (15.86–26.19)
Logical expansion	S3: S4 minus S1	Official 450	21.03 (15.86–26.19)
Logical expansion	S3: S4 minus S1	Community clusters	21.03 (15.90–26.15)
Logical expansion	S3: S4 minus S1	County clusters	21.03 (15.94–26.12)

Note. Point estimates and weights are identical across variance specifications within each classification. Earlier subset uses zero singleton contributions; its average-contribution alternative is archived in eTable 31. None is a full reconstruction of original PPS uncertainty.

eTable 41. Direct discordant-state outcome contrasts under alternative variance specifications

Classification	Outcome	Variance specification	S3 vs S2 OR (95% CI)	Holm P
Complete-domain	Mortality	Earlier subset	1.27 (0.70–2.33)	0.864
Complete-domain	Mortality	Eligible 449	1.27 (0.70–2.33)	0.864
Complete-domain	Mortality	Official 450	1.27 (0.70–2.33)	0.864
Complete-domain	Mortality	Community clusters	1.27 (0.70–2.32)	0.860
Complete-domain	Mortality	County clusters	1.27 (0.70–2.31)	0.853
Complete-domain	Survivor ADL	Earlier subset	1.40 (1.03–1.91)	0.093
Complete-domain	Survivor ADL	Eligible 449	1.40 (1.03–1.91)	0.095
Complete-domain	Survivor ADL	Official 450	1.40 (1.03–1.91)	0.095
Complete-domain	Survivor ADL	Community clusters	1.40 (1.03–1.92)	0.102
Complete-domain	Survivor ADL	County clusters	1.40 (1.01–1.94)	0.123
Complete-domain	Death or disability	Earlier subset	1.41 (1.05–1.90)	0.093
Complete-domain	Death or disability	Eligible 449	1.41 (1.05–1.90)	0.094
Complete-domain	Death or disability	Official 450	1.41 (1.05–1.90)	0.094
Complete-domain	Death or disability	Community clusters	1.41 (1.04–1.91)	0.102
Complete-domain	Death or disability	County clusters	1.41 (1.03–1.93)	0.123
Complete-domain	Survivor ADL + chronic count	Earlier subset	1.00 (0.69–1.44)	0.992
Complete-domain	Survivor ADL + chronic count	Eligible 449	1.00 (0.69–1.45)	0.992
Complete-domain	Survivor ADL + chronic count	Official 450	1.00 (0.69–1.45)	0.992
Complete-domain	Survivor ADL + chronic count	Community clusters	1.00 (0.69–1.45)	0.993
Complete-domain	Survivor ADL + chronic count	County clusters	1.00 (0.68–1.47)	0.993
Logical expansion	Mortality	Earlier subset	1.05 (0.62–1.78)	0.882
Logical expansion	Mortality	Eligible 449	1.05 (0.62–1.78)	0.884
Logical expansion	Mortality	Official 450	1.05 (0.62–1.78)	0.888
Logical expansion	Mortality	Community clusters	1.05 (0.63–1.75)	0.883
Logical expansion	Mortality	County clusters	1.05 (0.66–1.68)	0.880
Logical expansion	Survivor ADL	Earlier subset	1.19 (0.90–1.57)	0.882
Logical expansion	Survivor ADL	Eligible 449	1.19 (0.90–1.57)	0.884
Logical expansion	Survivor ADL	Official 450	1.19 (0.89–1.57)	0.888
Logical expansion	Survivor ADL	Community clusters	1.19 (0.90–1.57)	0.883
Logical expansion	Survivor ADL	County clusters	1.19 (0.89–1.57)	0.880
Logical expansion	Death or disability	Earlier subset	1.19 (0.90–1.56)	0.882
Logical expansion	Death or disability	Eligible 449	1.19 (0.90–1.56)	0.884
Logical expansion	Death or disability	Official 450	1.19 (0.90–1.57)	0.888
Logical expansion	Death or disability	Community clusters	1.19 (0.90–1.56)	0.883
Logical expansion	Death or disability	County clusters	1.19 (0.90–1.56)	0.880
Logical expansion	Survivor ADL + chronic count	Earlier subset	0.84 (0.60–1.17)	0.882
Logical expansion	Survivor ADL + chronic count	Eligible 449	0.84 (0.60–1.17)	0.884

Classification	Outcome	Variance specification	S3 vs S2 OR (95% CI)	Holm P
Logical expansion	Survivor ADL + chronic count	Official 450	0.84 (0.60–1.17)	0.888
Logical expansion	Survivor ADL + chronic count	Community clusters	0.84 (0.60–1.18)	0.883
Logical expansion	Survivor ADL + chronic count	County clusters	0.84 (0.59–1.19)	0.880

Note. Each classification/specification uses a four-test family. Confidence intervals are not multiplicity-adjusted. Chronic-count adjustment changes the comparison because chronic conditions contribute to the deficit index.

eTable 42. State associations under alternative clustering

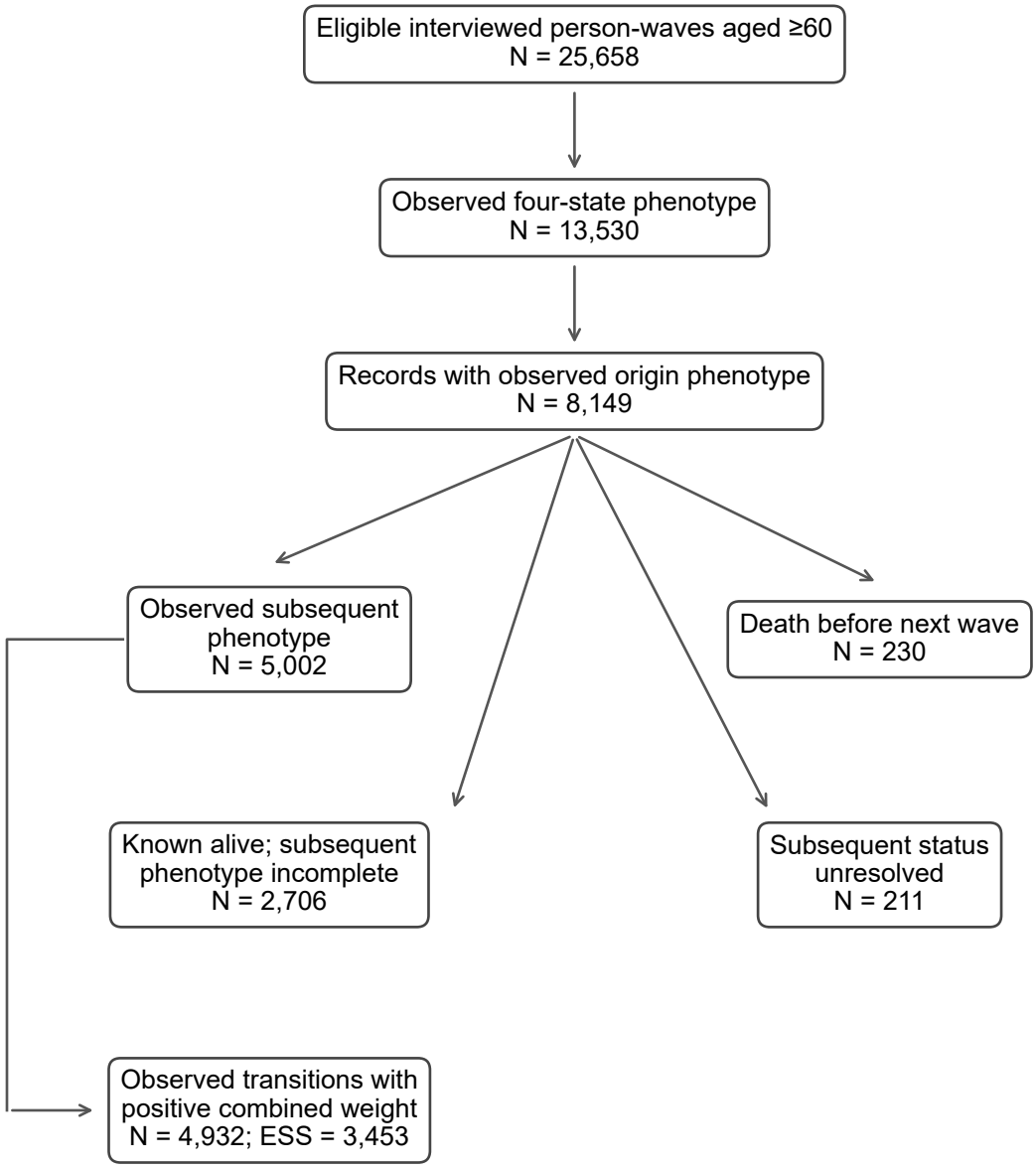
Classification	Outcome	State vs S1	Eligible 449 OR (95% CI)	Community clusters OR (95% CI)	County clusters OR (95% CI)
Complete-domain	Mortality	S2	0.99 (0.63–1.54)	0.99 (0.62–1.57)	0.99 (0.59–1.64)
Complete-domain	Mortality	S3	1.25 (0.76–2.06)	1.25 (0.77–2.04)	1.25 (0.78–2.02)
Complete-domain	Mortality	S4	2.32 (1.52–3.52)	2.32 (1.52–3.52)	2.32 (1.52–3.54)
Complete-domain	Survivor ADL	S2	1.73 (1.36–2.20)	1.73 (1.37–2.19)	1.73 (1.39–2.16)
Complete-domain	Survivor ADL	S3	2.43 (1.86–3.17)	2.43 (1.86–3.18)	2.43 (1.85–3.19)
Complete-domain	Survivor ADL	S4	3.73 (2.90–4.81)	3.73 (2.91–4.78)	3.73 (2.95–4.73)
Complete-domain	Death or disability	S2	1.60 (1.28–2.00)	1.60 (1.28–1.99)	1.60 (1.29–1.97)
Complete-domain	Death or disability	S3	2.25 (1.74–2.92)	2.25 (1.74–2.93)	2.25 (1.72–2.95)
Complete-domain	Death or disability	S4	3.53 (2.79–4.46)	3.53 (2.80–4.45)	3.53 (2.82–4.42)
Logical expansion	Mortality	S2	1.09 (0.76–1.56)	1.09 (0.76–1.57)	1.09 (0.75–1.59)
Logical expansion	Mortality	S3	1.14 (0.72–1.82)	1.14 (0.72–1.81)	1.14 (0.74–1.78)
Logical expansion	Mortality	S4	3.31 (2.42–4.53)	3.31 (2.41–4.55)	3.31 (2.38–4.62)
Logical expansion	Survivor ADL	S2	2.13 (1.74–2.61)	2.13 (1.75–2.60)	2.13 (1.77–2.57)
Logical expansion	Survivor ADL	S3	2.52 (1.98–3.22)	2.52 (1.97–3.24)	2.52 (1.95–3.27)
Logical expansion	Survivor ADL	S4	4.94 (4.07–6.00)	4.94 (4.07–5.99)	4.94 (4.08–5.98)
Logical expansion	Death or disability	S2	1.92 (1.58–2.34)	1.92 (1.59–2.32)	1.92 (1.60–2.30)
Logical expansion	Death or disability	S3	2.28 (1.80–2.89)	2.28 (1.79–2.90)	2.28 (1.77–2.94)
Logical expansion	Death or disability	S4	4.62 (3.85–5.54)	4.62 (3.85–5.54)	4.62 (3.84–5.56)

Note. Models, estimated weights and point estimates are fixed within each classification. The two unstratified specifications are sensitivity analyses, not conservative guarantees or clinical validation.

Supplemental References

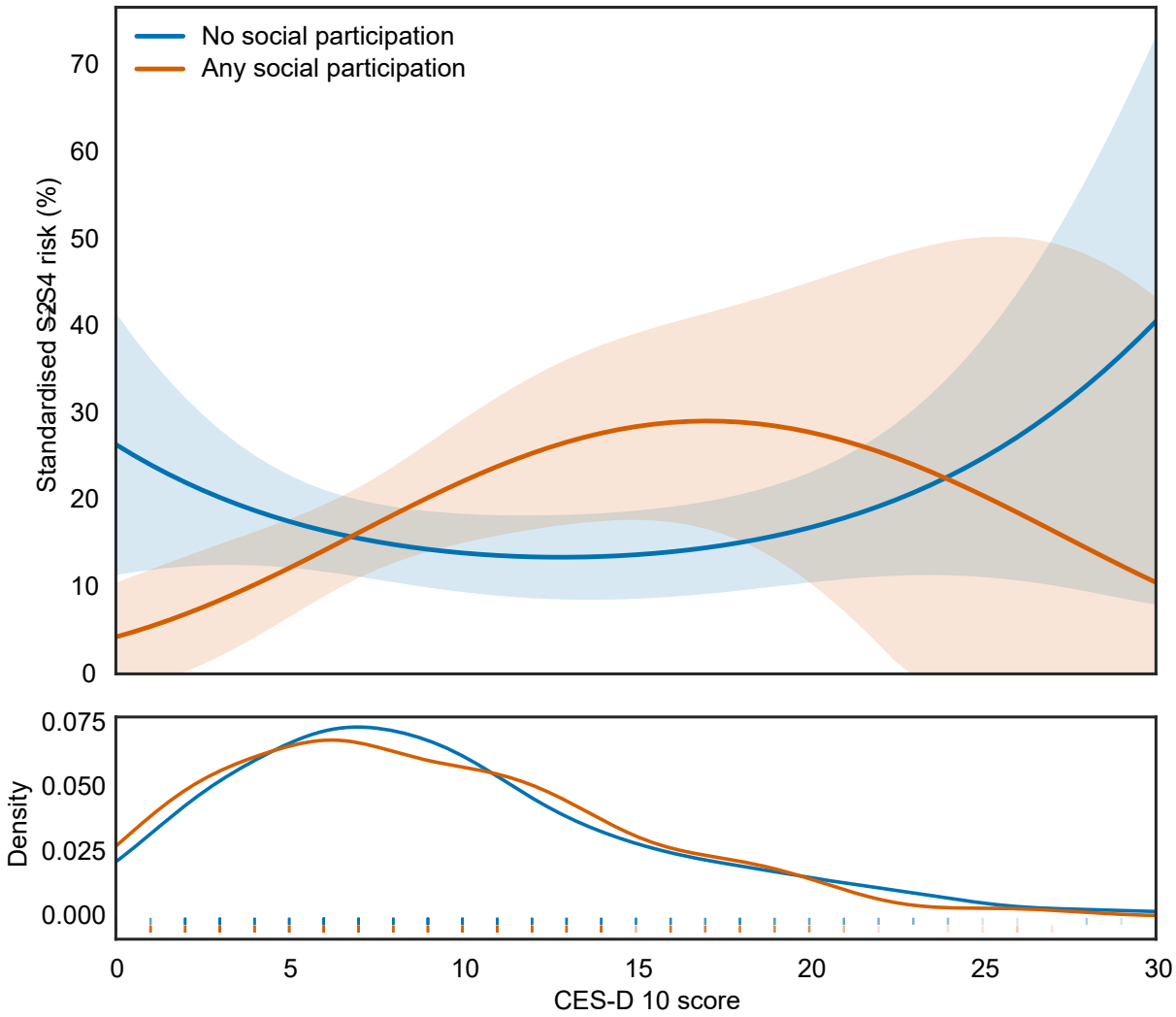
1. World Health Organization. Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. World Health Organization; 2017.
2. Chen LK, Woo J, Assantachai P, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc. 2020;21:300-307.e2. doi:10.1016/j.jamda.2019.12.012.
3. 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;20:465-470. doi:10.1002/gps.1314.

Supplementary Figures



eFigure 1. Participant and person-period flow. Boxes show eligible person-waves, observed phenotypes, origin records, observed destination phenotypes, deaths and the positive-weight analytic sample.

Alt text: A vertical flow diagram narrows from 25,658 eligible person-waves to 13,530 observed phenotypes, 8,149 origin records and 4,932 positive-weight observed transitions, with 230 deaths shown as an additional outcome.



eFigure 2. Exploratory quadratic depressive-symptom-by-social-participation model for S2-to-S4 progression. The upper panel shows marginally standardised risks with 95% confidence bands; the lower panel shows the observed CES-D distribution by participation group. The widening tails and non-monotonic fitted pattern indicate model sensitivity.

Alt text: Two fitted risk curves cross in the middle of the CES-D range and have wide confidence bands at both extremes; density curves show few observations at high scores.