

Attainment of 600 total MET-minutes per week and subsequent ADL disability or death among older adults with multimorbidity in China: a target-trial-informed longitudinal study

Gang Li

Harbin Institute of Technology

Qilin Chen

chenqilin_1024@163.com

Harbin Institute of Technology

Research Article

Keywords: physical activity, multimorbidity, activities of daily living, aging, target-trial framework, overlap weighting, CHARLS

Posted Date: September 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10806517/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

Older adults with multimorbidity face a high risk of losing independence in basic activities of daily living (ADLs). Static comparisons of active and inactive people do not directly address whether reaching an activity threshold after prior insufficient activity is associated with subsequent disability.

Methods

We conducted a target-trial–informed longitudinal analysis of the 2015, 2018, and 2020 waves of the China Health and Retirement Longitudinal Study. The 2018 interview was time zero; total physical activity below 600 metabolic equivalent of task minutes per week (MET-min/week) in 2015 was a historical eligibility condition. Participants were aged at least 60 years, had at least two reconstructed chronic conditions, were independent in all six basic ADLs, and had classifiable activity at both assessments. At time zero, they were classified as attaining or not attaining 600 total MET-min/week. The primary outcome was incident basic ADL disability or death by 2020. Propensity scores and overlap weights were re-estimated in each of 50 multiply imputed datasets to estimate the average treatment effect in the overlap population (ATO).

Results

The analysis included 271 participants (97 not attaining and 174 attaining the threshold), among whom 95 composite events occurred. Overlap-weighted risks were 0.420 and 0.314, respectively. The risk ratio was 0.748 (95% confidence interval [CI] 0.528 to 1.059), and the risk difference was -0.106 (95% CI -0.224 to 0.013), corresponding to an estimated 10.6 fewer events per 100 people in the ATO population. Mean overlap-weight effective sample size was 222.2. Fourteen otherwise eligible participants had indeterminate outcomes; prespecified extreme assignments retained risk-ratio point estimates below one. Results changed when activity was fully reconstructed as moderate-to-vigorous activity only.

Conclusions

Point estimates for attaining 600 total MET-min/week favored a lower two-year risk of ADL disability or death in the eligible overlap population, although the primary estimate was imprecise. This total-activity contrast includes walking and should not be interpreted as equivalent to attaining a moderate-to-vigorous activity guideline.

Background

Population aging has increased the number of people living for extended periods with more than one chronic condition. Basic activities of daily living (ADLs)—dressing, bathing, eating, transferring, toileting, and continence—represent essential self-care capacities. New difficulty in any of these tasks can alter living arrangements, increase dependence on relatives or paid carers, and accelerate use of health and long-term-care services. Longitudinal evidence from CHARLS and the Survey of Health, Ageing and Retirement in Europe suggests that

multimorbidity and ADL/IADL disability are strongly and bidirectionally related.[28] Preserving independence among older adults with multimorbidity is therefore a clinically important and public-health-relevant goal.

Physical activity is a plausible modifiable contributor to functional health. Regular movement may maintain lower-extremity strength, balance, cardiorespiratory capacity, joint mobility, and the ability to perform repeated everyday tasks. It may also support mood, cognition, sleep, and social engagement, which can influence whether physical capacity is translated into independent living. Systematic reviews generally associate higher physical activity with fewer functional limitations and more favorable ADL outcomes in older adults, although the strength of evidence varies by activity type, measurement method, and outcome definition.[2,3] Reviews focused on multimorbidity report that inactive adults with multiple conditions are common and that exercise programs can improve selected physical and psychosocial outcomes, but the trials are heterogeneous and often too small or short to establish effects on incident basic ADL disability.[4,5]

Longitudinal studies add a temporal dimension but answer different questions. In CHARLS, activity of at least 600 MET-min/week was associated with lower subsequent risk of a combined ADL/IADL impairment outcome in a broad middle-aged and older sample.[20] In the Toledo Study of Healthy Aging, persistently high activity and a trajectory that increased from a low level were associated with less incident or worsening disability than a low-decreasing trajectory.[21] Other cohorts have linked increased recreational activity, maintained or increased activity, and objectively measured total daily activity with later disability or mobility outcomes.[22–24] Among people with or at risk of knee osteoarthritis, increases from inactivity and greater light-intensity activity have also been associated with better disability trajectories.[25,26] These studies support attention to change, but differences in populations, exposure scales, and outcomes preclude treating them as estimates of one common intervention.

Randomized evidence is similarly nuanced. The LIFE trial showed that a structured program centered on moderate-intensity walking, with resistance, flexibility, and balance activities, reduced major mobility disability among sedentary older adults with functional limitations.[29] Dose analyses related greater achieved activity to better performance and less mobility disability.[31] Yet a secondary LIFE analysis did not show a reduction in basic ADL dependency or broader self-reported ADL/IADL disability, despite improvement in severe mobility disability.[30] An earlier trial in older adults with knee osteoarthritis reported lower incident ADL disability after aerobic or resistance exercise.[32] An unresolved issue is how to formulate the exposure. A static comparison of “active” with “inactive” participants mixes longstanding behavioral patterns with recent change and may be particularly vulnerable to confounding by health status. A more decision-relevant contrast asks what happens after a person who was previously insufficiently active reaches a prespecified threshold rather than remains below it. Even this contrast does not by itself identify an intervention: observed threshold crossing can reflect genuine behavior change, transient reporting variation, recovery from illness, or regression to the mean after selection on a low baseline value.[34] Self-reported physical activity has only limited agreement with direct measurement, and errors can occur in recalling frequency, assigning duration categories, or distinguishing intensity.[7,8,33]

The threshold also requires careful naming. IPAQ-style scoring assigns MET values to walking, moderate activity, and vigorous activity, then sums activity volume.[6] Consequently, 600 **total** MET-min/week may be reached partly or entirely through reported walking. It is not equivalent to the World Health Organization recommendation framed in moderate-to-vigorous minutes and muscle strengthening.[10] Published CHARLS work has used the

same 3.3, 4.0, and 8.0 MET values,[9,20,27] but operational conventions and routing patterns may materially change who crosses the threshold. Comparing total activity with a strict IPAQ classification and a fully rebuilt moderate-to-vigorous physical activity (MVPA) strategy is therefore a measurement-sensitivity exercise, not a test of an independently identified walking mechanism.

The distinction between status and attainment is substantive. A baseline comparison asks how people who report different activity levels subsequently differ. The present contrast begins with a common history of insufficient activity and asks whether reaching a defined total-volume boundary by time zero is associated with later function. That restriction makes the strategy legible while selecting a narrower population and exposing the design to threshold misclassification. Sánchez-Sánchez and colleagues showed why a low-increasing trajectory is not interchangeable with remaining low,[21] and Nguyen and colleagues used a target-trial emulation to align repeated activity strategies with follow-up.[37] Neither design estimates the present two-wave total-MET contrast.

We used a target-trial framework to align eligibility, initial strategy classification, time zero, follow-up, outcome, contrast, and analysis before interpreting results.[11,12] The framework reduces design ambiguity but cannot repair unmeasured confounding, exposure error, or inadequate temporal information. Our question was: among adults aged at least 60 years with multimorbidity, no basic ADL difficulty in 2018, and total activity below 600 MET-min/week in 2015, was attaining versus not attaining 600 total MET-min/week at the 2018 time zero associated with the two-year risk of incident basic ADL disability or death? We prespecified the composite as the primary outcome and targeted the ATO among eligible participants with common propensity-score support. The threshold describes observed total activity volume at time zero; it is not a fully specified intervention, clinical prescription, or sustained-adherence regimen.

Methods

Study design, setting, and data source

We conducted a target-trial–informed longitudinal analysis using the 2015, 2018, and 2020 waves of the China Health and Retirement Longitudinal Study (CHARLS). CHARLS is a longitudinal survey of community-dwelling Chinese adults aged 45 years or older and their spouses, selected through multistage probability sampling.[1] Health, demographic, household-economic, sample-status, and physical-measurement modules were linked by standardized person identifiers inside a controlled environment; no individual records or identifiers entered manuscript products.

Each wave had a distinct role. The 2015 survey established prior physical-activity insufficiency and provided lagged information. The 2018 interview was time zero, when clinical eligibility, covariates, and observed strategy were assessed. Follow-up ended at 2020 ADL ascertainment, death before or by that wave, or inability to determine either status. Thus, 2015 activity was a historical eligibility condition, not the start of follow-up.

The official 2015 questionnaire states that physical-activity items were administered to the main respondent and spouse in a randomly selected half of households.[19] We applied the questionnaire random-selection indicator before assessing item response. People outside that half-sample were classified as not sampled, never as inactive; item nonresponse and structural routing remained distinct.

Target trial specification and observational mapping

Table 1 maps the hypothetical protocol to the observational data. A corresponding trial would enroll eligible people at the 2018 time zero and assign an initial total activity-volume target of at least 600 MET-min/week versus remaining below it, with follow-up through 2020 or death. CHARLS supplied a single reported activity classification rather than a randomized, monitored intervention. The analysis modeled the probability of that classification from measured adjustment variables and targeted the ATO. Deviations included self-report, multiple ways of composing the same total MET value, no post-2018 adherence monitoring, contemporaneous measurement of some health variables, and incomplete outcome ascertainment.

The contrast is best understood as an initial classification analogue. It does not represent the effect of a unique exercise programme, a prescribed mixture of walking and higher-intensity activity, or sustained adherence after baseline. Post-2018 activity was unavailable, and no per-protocol analysis was attempted. A causal interpretation of the observational ATO would additionally require well-defined treatment versions sufficient for consistency, conditional exchangeability, positivity, correct imputation and propensity-score models, and correct outcome ascertainment.

Table 1
Hypothetical target trial and observational mapping

Component	Hypothetical target trial	CHARLS mapping	Principal deviation or identifying condition
Eligibility	People aged ≥ 60 years with ≥ 2 chronic conditions, independent in all six basic ADLs, and below 600 total MET-min/week at the historical assessment.	Age, reconstructed chronic-condition count and ADL independence were determined at the 2018 time zero; historical insufficiency and PA classifiability were determined from 2015 and 2018 responses.	The 2015 PA module was administered to a random half of households; self-report and routing limit measurement fidelity.
Initial activity strategies	At time zero, follow an activity strategy designed to attain ≥ 600 total MET-min/week versus remain below 600. A full trial would specify activity modes, delivery, monitoring and acceptable versions.	Reported 2018 walking, moderate and vigorous activity was converted to total MET-min/week; observed status defined the initial classification analogue.	A single observed classification is not an assigned, monitored or sustained programme; many activity compositions yield the same total.
Assignment	Random allocation at time zero.	No random allocation occurred. The probability of observed classification was modeled with prespecified adjustment variables, and overlap weighting targeted common support.	Conditional exchangeability is required and cannot be verified.
Time zero	The 2018 assessment, when eligibility would be confirmed and the initial strategy assigned.	The 2018 interview was the common time zero for eligibility, strategy classification, adjustment measurement and follow-up initiation.	Several health variables and recalled PA were measured in the same interview, making their within-wave order ambiguous.
Follow-up	From time zero until the 2020 assessment, death or outcome-ascertainment failure.	Outcome status was determined from 2020 ADL responses and mortality/exit information.	Interview spacing varied; 14 otherwise eligible participants had indeterminate outcomes.
Outcomes	Primary: incident basic ADL disability or death. Supporting: alive with incident ADL disability, death, and alive with preserved ADL independence.	Any difficulty in at least one of six ADLs or death by 2020 formed the composite. All supporting states used the full baseline denominator.	ADL is structurally unobserved after death; the three observed states do not identify latent ADL under survival. Death was sparse.
Contrast	Two-year risk under the two initial activity strategies, expressed as RR and RD.	Observed classification groups were compared after MI-specific propensity-score overlap weighting.	Classification does not ensure post-2018 adherence; the estimate is an observational analogue rather than a programme effect.
Estimand	ATO among eligible people whose measured characteristics support either initial strategy.	OW emphasized participants with greatest empirical propensity-score overlap.	The empirical ATO is neither the unweighted ATE for all eligible participants nor a nationally representative effect.

Component	Hypothetical target trial	CHARLS mapping	Principal deviation or identifying condition
Identification	Consistency, exchangeability under randomization, positivity, correct measurement and correct analysis.	Observational interpretation additionally requires conditional exchangeability, correct MI/PS models and adequate outcome ascertainment.	Unmeasured health change, exposure error, regression to the mean and informative loss remain possible.
Analysis	Estimate strategy-specific risks, RR and RD; address missing adjustment variables without imputing strategy or outcome.	Fifty MICE datasets; PS and OW refitted within each; joint sandwich variance and Rubin pooling; RR pooled on the log scale.	Survey weights were not used because the target was the empirical sample-overlap population. A community-cluster variance diagnostic was added without changing point estimates.

Notes: ADL, activity of daily living; ATO, average treatment effect in the overlap population; MICE, multiple imputation by chained equations; OW, overlap weighting; PA, physical activity; RD, risk difference; RR, risk ratio.

Eligibility

Eligibility required classifiable 2015 and 2018 activity; total activity below 600 MET-min/week in 2015; age at least 60 years; at least two chronic conditions; no difficulty in all six basic ADLs at time zero; and a determinable 2020 composite outcome. We did not use cleaned datasets that had removed deaths or nonrespondents before outcome construction.

Physical-activity measurement and strategies

Walking, moderate-intensity, and vigorous-intensity activity were reconstructed from participation, days per week, and duration category. The protocol mapped the four positive response bands—<30, 30 to < 120, 120 to < 240, and ≥ 240 minutes/day—to 20, 75, 180, and 240 minutes/day; explicit nonparticipation mapped to zero. These are analyst-defined representatives for categorical CHARLS responses, not values mandated by IPAQ. MET coefficients were 3.3, 4.0, and 8.0, respectively. Weekly activity was **total MET-min/week = $\Sigma(\text{MET} \times \text{minutes/day} \times \text{days/week})$** . A band contributed zero only after an explicit report of no participation. Structural skips, unresolved routes, and invalid day or duration values were not zero.

Total activity below 600 MET-min/week was insufficient. Strategies compared participants below 600 in 2015 who remained below 600 in 2018 with those who reached at least 600 in 2018. Because walking contributes to the sum, this is not the WHO MVPA recommendation.[10] A strict-IPAQ sensitivity also required applicable day conditions. A fully rebuilt MVPA analysis excluded walking and reconstructed 2015 eligibility, 2018 strategy, sample, propensity scores, weights, and ESS; a second MVPA analysis reclassified strategy within the fixed primary population.

Chronic-condition reconstruction

The primary count integrated wave-specific diagnoses, preloaded histories, routing, prior states, and explicit corrections. Each disease-wave state was *known yes*, *known no*, *carried forward*, *corrected no*, or *unknown*. Prior explicit diagnoses were carried forward unless later corrected, and structural skips were not coded as absence. Multimorbidity required at least two known-yes or carried-forward conditions. Prespecified checks used Harmonized CHARLS counts and the eligibility intersection; the raw reconstruction remained primary independent of effect direction.

Outcomes

The primary outcome was incident basic ADL disability or death by 2020. The six ADLs were dressing, bathing, eating, transferring, toileting, and continence. Wave-specific value labels defined no difficulty, difficulty without help, need for help, or inability. Baseline independence required six no-difficulty responses; any later difficulty among living respondents defined incident disability. Death was a composite event, post-death ADL was structurally nonexistent and not imputed, and the composite was unknown when neither ADL nor death could be determined.

Supporting states were alive with incident ADL disability, dead, and alive with preserved ADL independence. All used the full baseline denominator, were mutually exclusive and exhaustive, and summed to one after weighting. The first is not survivor-conditioned. Death nevertheless truncates a different latent ADL-under-survival question, which we did not estimate.[16,35]

Covariates

The propensity-score set comprised age, sex, marital status, education, log household consumption per capita, urban/town residence, current smoking, past-year alcohol use, chronic-condition count, cognition, CES-D-10 symptoms, body pain, IADL-limitation count, 2015 total activity, social participation, and most recent pre-2018 BMI. The main logistic propensity model therefore contained 16 covariate terms plus an intercept. Selection was based on subject-matter relevance and availability by the 2018 interview, not statistical significance. Residence was distinct from household registration; missing 2018 residence used the nearest valid earlier residence.

Several health measures, including pain and IADL, were collected in the same 2018 interview as physical activity. We treated them as prespecified time-zero adjustment variables because they can influence both activity classification and subsequent function, while recognizing that their temporal order relative to recalled activity is ambiguous. A structural sensitivity replaced current measures with primarily 2015 values where available. Additional file 1 (Figure S3) displays the resulting temporal assumptions, the possible influence of health change between waves, and the role of self-report and threshold error.

Missing data

Treatment, composite outcome, death, known ADL state, and structurally nonexistent post-death ADL were not imputed. Baseline adjustment variables were imputed by chained equations in 50 datasets, with treatment, outcome, earlier-wave measures, and other available auxiliary variables included as predictors. Continuous variables preferentially used predictive mean matching; binary and unordered categorical variables used logistic

or multinomial models. Derived scores were reconstructed passively. Before seeded imputation, records were sorted by standardized identifier to support reproducibility.

We assessed traces, observed and imputed distributions, ranges, passive constraints, fraction of missing information (FMI), and Monte Carlo error. Propensity scores and weights were re-estimated within every completed dataset.[15] Complete-case OW and a model omitting cognition were sensitivity analyses. Unknown outcomes were not imputed; four deterministic extreme assignments assessed their potential influence. Multiple imputation represented uncertainty in missing covariates and did not convert all 271 records into fully observed cases.

Propensity scores and overlap weighting

Within each imputation, logistic regression estimated $e(X)$, the probability of attaining the threshold. Attainers received weight $1 - e(X)$ and non-attainers received weight $e(X)$. [13,14] This overlap weighting (OW) targets eligible participants with the greatest empirical support for either strategy, rather than all eligible CHARLS respondents or all Chinese older adults with multimorbidity. Diagnostics included propensity-score distributions and tails, model convergence and separation, maximum weights, standardized mean differences (SMDs), and effective sample size (ESS), calculated as $(\sum w)^2 / \sum w^2$ overall and by strategy. Near-zero fitted-term SMDs were interpreted alongside score support and weight concentration because logistic-PS OW has an exact mean-balance property. Stabilized inverse-probability-of-treatment weighting (IPTW), before and after first/99th-percentile truncation, was a sensitivity analysis and targets a different weighted population.

Effect estimation and uncertainty

Within each imputation, we estimated strategy-specific weighted risks, attained-versus-not-attained risk ratios (RRs), and risk differences (RDs). Joint sandwich variance incorporated the estimated propensity-score parameters and both weighted outcome means. RD was analysed on the identity scale and RR on the log scale. Rubin's rules combined estimates across the 50 imputations; pooled log RRs were exponentiated.

All supporting-state indicators used the complete 271-person baseline denominator, so the weighted probabilities of being alive with incident ADL disability, dead, and alive with preserved ADL independence summed to one within each strategy. A 500-replicate participant bootstrap in one imputation audited the within-imputation sandwich variance; it did not include between-imputation uncertainty and did not replace the Rubin-pooled primary confidence interval.

Survey design and dependence

The primary analysis did not apply CHARLS survey weights because the estimand was the empirical overlap-population effect within the narrowly eligible 2015 activity half-sample, not a nationally standardized effect. The primary sandwich used participant-level score contributions. An additional, non-prespecified variance sensitivity analysis summed those contributions within 2018 communities without changing the imputed data, propensity scores, OW, or point estimates. The 271 participants represented 177 communities.

Prespecified sensitivity analyses

Prespecified analyses examined missing-data and weighting choices (complete-case OW; stabilized IPTW), chronic-condition definitions (Harmonized counts and the definition intersection), activity definitions (strict IPAQ

day rules, fully rebuilt MVPA, fixed-population MVPA, and alternative duration representatives specified before effect estimation), covariate timing and specification (primarily 2015 measures, cognition omission, and a flexible propensity-score form), pooled windows 1 and 3, and four assignments for the 14 unknown outcomes. Analyses that changed the activity definition, eligibility population, weights, or trial windows were labelled as different estimands rather than treated as direct replications of the primary ATO. The first window was not estimated separately because of poor propensity-score support; the second was excluded before formal effect estimation because one strategy group contained 19 participants and eight composite events.

Identification and reporting

Causal interpretation requires consistency, conditional exchangeability, positivity, correct imputation and propensity-score specifications, and adequate outcome ascertainment. These assumptions are not fully verifiable; residual confounding, reverse causation, contemporaneous timing, regression to the mean, and activity misclassification remain possible. The target-trial framework structures the question but cannot repair these data limitations.[11] Reporting followed TARGET and STROBE.[12,17] No post hoc subgroup, threshold, or outcome redefinition was introduced.

Use of generative artificial intelligence in manuscript preparation

OpenAI Codex was used during manuscript preparation for English-language editing, document formatting, and organizational assistance. It was not used to determine the research question, define the analysis population, exposure, outcomes, estimand, or statistical model, perform or alter statistical analyses, or make unreviewed interpretations. The authors verified all numerical results, references, and submitted text and take full responsibility for the work.

Results

Participant selection

Figure 1 shows the selection sequence. The starting denominator was 19,752 unique 2018 health-module respondents, of whom 18,044 linked to the 2015 health module. The 2015 household random-half selection retained 9,063; 8,658 had at least one relevant activity response, and 8,396 had classifiable 2015 activity. Classifiable activity in both 2015 and 2018 was available for 8,391. Sequential application of the age, multimorbidity, time-zero ADL-independence, and prior-insufficiency requirements retained 4,713, 3,016, 1,596, and 285 participants, respectively.

Of the 285 people meeting every pre-outcome criterion, 14 (4.9%) had neither determinable 2020 ADL status nor confirmed death. These comprised 6 in the not-attained group and 8 in the attained group. The primary analysis therefore included 271 participants: 97 who remained below 600 total MET-min/week and 174 who attained the threshold. The reduction from the source cohort was driven principally by the 2015 PA household subsample and the conjunction of eligibility conditions, not by propensity-score trimming or complete-case restriction (Fig. 1).

Baseline characteristics

Before weighting, participants who attained the threshold were younger on average and more often married or partnered, reported alcohol use, and participated socially; they also had fewer IADL limitations. Table 2 presents observed-case unweighted means, standard deviations, counts and percentages, together with missingness and the average overlap-weighted summaries across imputations. Unweighted and weighted SMDs were calculated within each completed dataset and averaged. These weighted summaries characterize the overlap population and are not participant counts.

The largest unweighted difference was age (absolute SMD 0.450): means were 72.85 and 69.44 years in the not-attained and attained groups. IADL limitation means were 1.02 and 0.66 (absolute SMD 0.285). Marital status and social participation also differed. Mean 2015 total activity remained low in both groups, 134.22 and 157.53 MET-min/week, reflecting their shared historical eligibility. The weighted overlap population had a mean age of 71.07 years, was 60.8% female, had a mean 3.54 chronic conditions and 0.86 IADL limitations, and had a mean imputed cognition score of 11.04. These are weighted covariate summaries, not the characteristics of 222 selected individuals.

Table 2
Baseline characteristics before and after overlap weighting

Characteristic	Not attained, unweighted	Attained, unweighted	Missing, n (%)	Absolute SMD, unweighted	Not attained, OW	Attained, OW	Absolute SMD, OW
Age, years	72.85 (8.36)	69.44 (6.76)	0 (0.0%)	0.450	71.07 (7.48)	71.07 (7.32)	0.000
Female	62/97 (63.9%)	103/174 (59.2%)	0 (0.0%)	0.097	60.8%	60.8%	0.000
Married or partnered	68/97 (70.1%)	140/174 (80.5%)	0 (0.0%)	0.242	76.1%	76.1%	0.000
Less than lower secondary	90/97 (92.8%)	167/174 (96.0%)	0 (0.0%)	0.139	94.5%	94.4%	0.007
Upper secondary or vocational training	6/97 (6.2%)	7/174 (4.0%)	0 (0.0%)	0.098	5.3%	5.6%	0.014
Tertiary education	1/97 (1.0%)	0/174 (0.0%)	0 (0.0%)	0.144	0.2%	0.0%	0.056
Log household consumption per capita	9.06 (1.46)	9.16 (1.01)	49 (18.1%)	0.084	9.12 (1.21)	9.12 (1.09)	0.000
Urban or town residence	11/97 (11.3%)	25/173 (14.5%)	1 (0.4%)	0.100	13.0%	13.0%	0.000
Current smoking	23/97 (23.7%)	32/174 (18.4%)	0 (0.0%)	0.131	22.2%	22.2%	0.000
Alcohol use in the past year	14/97 (14.4%)	38/174 (21.8%)	0 (0.0%)	0.193	17.3%	17.3%	0.000
Number of chronic conditions	3.65 (1.97)	3.52 (1.62)	0 (0.0%)	0.074	3.54 (1.86)	3.54 (1.70)	0.000
Cognition score	13.98 (5.19)	14.33 (5.17)	139 (51.3%)	0.208	11.04 (5.86)	11.04 (5.54)	0.000
CES-D-10 score	9.63 (6.35)	9.27 (6.40)	27 (10.0%)	0.069	9.70 (6.31)	9.70 (6.45)	0.000
Body pain	62/97 (63.9%)	121/174 (69.5%)	0 (0.0%)	0.120	65.8%	65.8%	0.000
Number of IADL limitations	1.02 (1.36)	0.66 (1.17)	0 (0.0%)	0.285	0.86 (1.24)	0.86 (1.31)	0.000
2015 total PA, MET-min/week	134.22 (197.96)	157.53 (215.07)	0 (0.0%)	0.113	143.97 (199.33)	143.97 (208.40)	0.000

Characteristic	Not attained, unweighted	Attained, unweighted	Missing, n (%)	Absolute SMD, unweighted	Not attained, OW	Attained, OW	Absolute SMD, OW
Social participation	30/97 (30.9%)	73/174 (42.0%)	0 (0.0%)	0.231	35.5%	35.5%	0.000
Pre-2018 BMI, kg/m ²	23.63 (4.33)	23.99 (3.59)	11 (4.1%)	0.103	23.77 (4.33)	23.77 (3.64)	0.000

Notes: Unweighted continuous values are observed-case mean (SD); categorical values are n/observed N (%). OW values are mean weighted summaries across 50 imputations. Missingness refers to the 271-person analytic sample. SMDs are absolute and averaged across completed datasets.

Missing data and weighting diagnostics

Cognition had the most missing baseline information (139/271, 51.3%), followed by household consumption (49/271, 18.1%), depressive symptoms (27/271, 10.0%), prior BMI (11/271, 4.1%), and residence (1/271, 0.4%). Requiring all adjustment variables to be observed left 108 complete cases; this was the sample for a sensitivity analysis, not the primary analysis. Tables S1–S3 and Figure S1 report missingness, observed-versus-imputed distributions, traces, ranges, and numerical diagnostics.

No completed dataset contained an out-of-range value or passive-constraint violation. Cognition-component trace sequences showed no sustained trend. An internal, non-prespecified 0.10 standardized-drift warning was exceeded for three components (0.128, 0.166, and 0.174); review of the full traces, distributions, constraints, Monte Carlo error, FMI, cognition-omission analysis, and exact reproduction of the primary estimates found no anomaly with a material effect on those estimates. RD FMI was 0.0053 and Monte Carlo error was 0.0006 (1.02% of its pooled standard error); log(RR) FMI was 0.0046 and Monte Carlo error was 0.0017 (0.95%).

Across 50 completed datasets, mean overall OW ESS was 222.2; mean strategy-specific ESS values were 88.5 for not attaining and 149.1 for attaining. All propensity-score models converged. Scores ranged from 0.053 to 0.954, no score was below 0.05, and at most one exceeded 0.95 in an imputation. The largest OW was 0.910. Figure 2 and Table S4 show common support, balance, and weight concentration across the multiple-imputation analysis.

Primary composite outcome

There were 95 unweighted composite events: 42 among the 97 participants who did not attain the threshold and 53 among the 174 who attained it. The corresponding raw counts are descriptive and are not the adjusted risks. After MI-specific propensity-score estimation and overlap weighting, estimated two-year risks of incident basic ADL disability or death were 0.420 for not attaining and 0.314 for attaining 600 total MET-min/week. The ATO RR was 0.748 (0.528 to 1.059), and the ATO RD was – 0.106 (–0.224 to 0.013) (Table 3). Both primary confidence intervals included the null.

On the absolute scale, the point estimate corresponds to 10.6 fewer composite events per 100 people in the ATO population. Its 95% CI was compatible with 22.4 fewer to 1.3 more events per 100. The relative and absolute scales therefore tell the same story: the central estimate favored attainment, while the data also permit little or no difference.

Table 3
Primary outcome and mutually exclusive outcome-state estimates

Outcome	Not attained, events/N	Attained, events/N	Not attained weighted risk	Attained weighted risk	Risk ratio (95% CI)	Risk difference (95% CI)
Incident basic ADL disability or death (primary)	42/97	53/174	0.420	0.314	0.748 (0.528 to 1.059)	-0.106 (-0.224 to 0.013)
Alive with incident basic ADL disability (supporting state)	38/97	42/174	0.391	0.230	0.590 (0.375 to 0.927)	-0.160 (-0.276 to -0.044)
Death (supporting state)	4/97	11/174	0.029	0.084	2.879 (0.572 to 14.494)	0.055 (0.002 to 0.107)
Alive and basic-ADL independent (supporting state)	55/97	121/174	0.580	0.686	1.182 (0.983 to 1.422)	0.106 (-0.013 to 0.224)

Notes: Events/N are unweighted descriptive counts; risks, RR and RD are MI/OW estimates for the ATO. Supporting states use the full baseline denominator and do not replace the primary composite.

Outcome-state and component results

The unweighted composite comprised 80 living participants with incident basic ADL disability and 15 deaths. The raw state counts were 38/97 versus 42/174 for alive with incident ADL disability, 4/97 versus 11/174 for death, and 55/97 versus 121/174 for alive and ADL-independent. These counts are descriptive and remain distinct from the weighted ATO risks in Table 3. For the unconditional state “alive with incident ADL disability,” weighted risks were 0.391 and 0.230; the RR was 0.590 (0.375 to 0.927) and the RD was – 0.160 (–0.276 to –0.044). For death, weighted risks were 0.029 and 0.084; the RR was 2.879 (0.572 to 14.494) and the RD was 0.055 (0.002 to 0.107). The death point estimate was in the opposite direction from the ADL state, and its ratio interval was extremely wide because deaths were sparse.

The weighted probabilities of being alive and remaining basic-ADL independent were 0.580 and 0.686, with RR 1.182 (0.983 to 1.422) and RD 0.106 (–0.013 to 0.224). Within each strategy, the probabilities of alive with disability, death, and alive independent summed to one. The composite pattern was numerically driven mainly by the alive-with-incident-ADL state, but that state remained supportive rather than primary. Its interval cannot supersede the prespecified composite, and the opposite death direction precludes summarizing the components as uniformly favorable.

Figure 3. Overlap-weighted composite risks and mutually exclusive outcome-state probabilities. State probabilities use the full baseline denominator and sum to one within strategy. The figure shows how the composite is distributed between alive with incident ADL disability and death. ADL, activity of daily living; OW, overlap weighting.

Sensitivity analyses

Table 4 and Fig. 4 group sensitivity analyses by their relation to the primary question. Analyses that retained the total-activity strategy gave point estimates near the primary RR. Complete-case OW produced RR 0.769 (95% CI 0.396 to 1.493). Stabilized IPTW yielded RR 0.749 (0.521 to 1.079) before truncation and 0.737 (0.511 to 1.063) after truncation. The Harmonized chronic-condition definition gave RR 0.802 (0.571 to 1.127). Models using primarily 2015 covariates, omitting cognition, and allowing a flexible propensity-score form gave RRs of 0.788, 0.753, and 0.721, respectively. The additional community-clustered variance analysis left the point estimate unchanged and modestly widened the RR interval to 0.523–1.070.

The four deterministic assignments for 14 unknown outcomes retained all 285 otherwise eligible participants and produced point RRs from 0.670 to 0.852 (Tables S5 and S6). These scenarios preserved the point direction but do not describe every possible informative-loss mechanism.

Changing the activity construct had a larger influence. Strict IPAQ classification with day conditions yielded RR 0.764 (0.562 to 1.040). The fully rebuilt MVPA-only analysis redefined historical insufficiency, strategy, eligibility, propensity scores, weights, and the overlap population, and yielded RR 1.168 (0.910 to 1.500). MVPA classification within the fixed primary population gave RR 0.978 (0.683 to 1.400). These estimates address different strategies or target populations and do not identify an independent effect of walking. Table S7 provides group counts, weighted risks, ESS, and specification changes for the complete set of analyses; Fig. 4 distinguishes approximately comparable checks from analyses that changed the estimand.

Table 4
Sensitivity analyses organized by relation to the primary estimand

Analysis	Relation to primary estimand	N	Events	RR (95% CI)	RD (95% CI)	Activity definition changed
Primary MI overlap weighting	primary estimand	271	95	0.748 (0.528 to 1.059)	-0.106 (-0.224 to 0.013)	No
Community-clustered variance diagnostic	Same point estimand; variance clustered by 2018 community	271	95	0.748 (0.523 to 1.070)	-0.106 (-0.227 to 0.015)	No
Complete-case overlap weighting	approximately same strategy; complete-case population	108	33	0.769 (0.396 to 1.493)	-0.077 (-0.259 to 0.106)	No
Stabilized IPTW, P1/P99 truncated	different weighting estimand	271	95	0.737 (0.511 to 1.063)	-0.109 (-0.229 to 0.011)	No
Harmonized chronic-condition definition	different eligibility definition and overlap population	274	96	0.802 (0.571 to 1.127)	-0.081 (-0.200 to 0.038)	No
Primarily 2015 lagged covariates	same sample/strategy; covariate timing changed	271	95	0.788 (0.564 to 1.101)	-0.088 (-0.206 to 0.031)	No
Cognition omitted from PS	same sample/strategy; PS covariates changed	271	95	0.753 (0.531 to 1.067)	-0.103 (-0.223 to 0.016)	No
Flexible propensity-score form	same sample/strategy; PS functional form changed	271	95	0.721 (0.506 to 1.027)	-0.117 (-0.234 to 0.000)	No
Strict IPAQ day-rule reconstruction	different PA strategy and eligible overlap population	321	111	0.764 (0.562 to 1.040)	-0.095 (-0.197 to 0.008)	Yes
MVPA within fixed primary population	same participants; different strategy and overlap target	271	95	0.978 (0.683 to 1.400)	-0.008 (-0.129 to 0.114)	Yes
Fully rebuilt MVPA-only strategy	different PA strategy and eligible overlap population	672	212	1.168 (0.910 to 1.500)	0.046 (-0.026 to 0.117)	Yes
Unknown outcomes: worst for attained	same pre-outcome cohort; deterministic outcome assignment	285	103	0.852 (0.616 to 1.180)	-0.059 (-0.175 to 0.058)	No
Unknown outcomes: best for attained	same pre-outcome cohort; deterministic outcome assignment	285	101	0.670 (0.473 to 0.951)	-0.149 (-0.266 to -0.032)	No

Notes: Different rows need not estimate the same population contrast. Full group counts, weighted risks, ESS and specification fields are in Table S7.

Computational and variance audit

Canonical sorting before seeded imputation reproduced every primary estimate and confidence limit exactly. In the single-imputation variance check, all 500 bootstrap replicates succeeded; bootstrap and joint-sandwich standard errors were similar for log(RR) (0.1768 versus 0.1762) and RD (0.0661 versus 0.0607). The bootstrap did not replace the Rubin-pooled primary intervals.

Discussion

Principal findings

Among ADL-independent older adults with multimorbidity and total activity below 600 MET-min/week in 2015, attaining that threshold at the 2018 time zero was associated with a lower estimated two-year risk of ADL disability or death in the overlap population. The RR was 0.748 and the RD corresponded to 10.6 fewer composite outcomes per 100 people. Their confidence intervals also accommodated little or no difference, so the result supports a lower-risk direction without establishing a definitive reduction.

The composite pattern was driven numerically by the state of being alive with incident ADL disability. Death was uncommon, pointed in the opposite direction, and had a wide ratio interval. Reporting three mutually exclusive states with the full baseline denominator avoided coding death as no disability or conditioning the ADL analysis on survival. The component estimates nonetheless cannot replace the prespecified composite or establish a mortality difference.

Why attainment differs from a static comparison

Requiring low total activity in 2015 and classifying strategies at the 2018 time zero defines a transition question rather than a static contrast. Requiring low total activity in 2015 sets a common historical condition. Classifying strategies at the 2018 time zero separates that history from follow-up. People may cross the boundary because symptoms improve, work or caregiving changes, reporting fluctuates or purposeful activity increases. Alignment makes these possibilities and the exchangeability assumption visible; it does not remove them. Nguyen et al. followed repeated activity strategies in more than 11,000 Australian women,[37] whereas Toledo identified empirical PASE trajectories.[21] Both support attention to activity history, but neither estimates this two-wave threshold ATO.

Selection on a single 2015 activity report also creates scope for regression to the mean. Random upward variation can move a participant across 600 MET-min/week in 2018 even without a sustained behavioral change, and the categorical duration scale can amplify threshold crossing. Mean 2015 activity was low in both strategies, but this does not distinguish durable change from transitory fluctuation. The strict-IPAQ analysis retained the point direction, and the primarily lagged-covariate model gave RR 0.788, yet neither test eliminates regression to the mean. Repeated pre-time-zero activity measurements would be required to distinguish sustained insufficiency from one low report.

Comparison with longitudinal cohort studies

A broad CHARLS cohort associated baseline activity of at least 600 MET-min/week with less subsequent combined ADL/IADL impairment.[20] Unlike that static comparison, our study required prior insufficiency, multimorbidity, and time-zero ADL independence. The Toledo cohort found less incident disability in a low-increasing than a low-decreasing activity trajectory,[21] while InCHIANTI related maintained or increased activity to later 400-m mobility.[24] Other cohorts linked increased recreational activity or actigraphy-measured total activity with less later disability.[22,23]

Studies in people with or at risk of knee osteoarthritis associated increases from objective inactivity and greater light-intensity time with better disability outcomes.[25,26] Their accelerometer metrics, disability scales and clinical populations differ from CHARLS. Multimorbidity further complicates comparison because symptoms can restrict activity, and multimorbidity and disability are bidirectionally associated.[28]

Objective-measure cohorts clarify a separate boundary. Shah et al. followed 584 initially ADL-independent older adults with actigraphy and reported a hazard ratio near 0.75 per 100,000 activity counts/day for incident disability, translating representative activity differences to an absolute scale.[23] Dunlop et al. used accelerometry in adults with or at risk of knee arthritis and observed lower incident disability across higher light-activity quartiles after accounting for MVPA.[26] These studies support total or lower-intensity movement as legitimate constructs, but they do not validate a self-reported 600-MET boundary or identify walking as the source of our estimate. Martínez-Hernández et al. used CHAMPS with Barthel, Lawton and performance outcomes in Mexican older adults,[38] while Sharif et al. modeled three-year transition from independence to ADL limitation in Malaysia.[39] Their instruments, populations, endpoints and handling of follow-up differ from the present ATO.

Comparison with exercise trials and systematic reviews

The LIFE randomized trial tested a structured moderate-intensity walking, resistance, flexibility, and balance program in sedentary adults with functional limitations and reduced objectively defined major mobility disability.[29] Achieved activity dose related to better performance and less mobility disability.[31] However, LIFE did not reduce basic ADL dependency or broad self-reported ADL/IADL disability, despite less severe mobility disability.[30] An osteoarthritis trial reported fewer incident ADL limitations after aerobic or resistance exercise.[32] Reviews likewise show favorable average patterns with heterogeneity in intervention and outcome.[3,5]

Our observed total-MET status was neither assigned nor supervised, and maintenance after 2018 was unknown. It should not be equated with these intervention programs or their performance-based outcomes.

Total PA, strict IPAQ, and MVPA

The primary exposure included walking at 3.3 METs. Strict IPAQ day rules retained a similar point direction, whereas the fully rebuilt MVPA analysis changed direction and fixed-population MVPA was close to the null. These analyses do not estimate one interchangeable intervention: a full rebuild changes historical eligibility, observed strategy, propensity-score support, and the overlap population, while the fixed-population version changes the strategy within the original cohort.

Objective-measure cohorts show that total and lower-intensity movement are meaningful activity constructs in other populations,[23,26] but the present data do not isolate walking from other contributors to total MET-

min/week. Self-report, categorical duration representatives, intensity classification, and regression to the mean after selection on low 2015 activity can all affect threshold crossing.[7,8,33,34] The divergence therefore identifies activity-definition sensitivity and limits translation of the primary total-activity estimate to an MVPA-only recommendation.

Composite outcome, ADL state, and death

The composite avoided coding death as no disability or deleting decedents. Because disability and death differ in importance, frequency, and point direction, component reporting is essential.[36] The three-state probabilities describe observed outcomes from the baseline cohort. They do not solve truncation by death for a latent disability-under-survival question.[16,35]

The death RR and RD differed in null coverage because RR was pooled on the log scale and RD on the identity scale while the reference death risk was very small. The formulas and scale-specific pooling reproduced correctly. With only 15 deaths, neither scale supports a strong mortality interpretation, and the result is not safety evidence.

Implications for geriatric care and healthy ageing

If supported by larger studies, the transition framing could inform trials aimed at older adults with multimorbidity who remain independent but report insufficient activity. The 600 total MET-min/week boundary is not a prescription: identical totals can arise from transport walking, occupational or household activity, and deliberate exercise. A trial would need to define activity modes and progression, adapt delivery to pain, falls risk and cardiovascular limitations, monitor adherence objectively, and measure function as well as activity volume. The present estimate identifies a population and decision point for that work; it does not specify an intervention package or replace individualized assessment.

ATO population and transportability

OW emphasized people whose measured covariates supported either observed strategy, limiting extrapolation and avoiding highly concentrated weights.[13,14] The mean overall ESS of 222.2 indicated moderate rather than severe weight concentration; it was not a count of retained participants. The empirical ATO applies to the eligible overlap population drawn from the 2015 random household activity half-sample. It is neither a nationally representative effect nor an ATE for all Chinese older adults with multimorbidity.

Within this population, the RD point estimate is clinically interpretable: 10.6 fewer composite outcomes per 100 people, with a 95% CI ranging from 22.4 fewer to 1.3 more. That uncertainty supports replication rather than a clinical prescription. A trial would also need to specify activity modes, delivery, safety adaptation, and post-assignment adherence, none of which is encoded by a single total-MET classification.

Strengths and limitations

The three-wave design separates prior insufficiency, strategy classification and follow-up, while the household sampling indicator prevents people who were not asked PA questions from being labeled inactive. Disease histories were reconstructed with routing-aware states, deaths were retained, supporting-state denominators were audited, and every imputation received a new PS/OW fit. Absolute risks, RR, RD, ESS, balance, FMI, Monte

Carlo error, deterministic outcome assignments and the community-clustered variance diagnostic provide complementary checks on the estimate.

Self-reported categorical PA remains the main measurement boundary. Recall, social-desirability error, intensity interpretation and analyst-defined duration representatives can move participants across 600 MET-min/week; selection below the threshold in 2015 also permits regression to the mean. The direction of resulting bias is not predictable because reporting error may differ by health and strategy. Strict IPAQ and MVPA analyses demonstrate that operational definitions matter but cannot recover an error-free exposure. Repeated accelerometry before and after time zero would address this more directly.

Exchangeability remains unverified. Some health adjustment variables and PA were measured at the same 2018 interview, and disease severity, frailty, recent symptom change, access to safe activity or motivation may remain incompletely measured. These variables can be proxies for pre-existing confounding but may also reflect health changes close to activity classification, creating temporal ambiguity and possible overadjustment. The primarily 2015 covariate model retained the point direction, with RR 0.788 (0.564 to 1.101), but no measured-data model eliminates reverse causation or residual confounding.

Cognition was missing for 51.3% of participants. MI included the known-outcome cohort using observed and auxiliary information; valid inference still depends on the imputation model and a conditionally ignorable missingness mechanism. Distribution and chain diagnostics found no out-of-range or logical violations, and omitting cognition produced RR 0.753 rather than materially shifting the point estimate. Fourteen outcomes were unknown, death involved 15 events and conjunctive eligibility produced a narrow cohort. Extreme assignments retained RRs below one but do not represent every informative-loss mechanism.

The analytic sample arose from the 2015 PA random-half module and several conjunctive eligibility conditions. Its size is therefore a consequence of the target question rather than propensity-score trimming, but it limits precision, especially for mortality. Survey weights were not used because the estimand was the empirical sample-overlap ATO; the result must not be interpreted as a nationally representative effect. Community-clustered standard errors changed the interval little, but residual household, regional and sampling-design dependence cannot be ruled out. The estimate is most applicable to community-dwelling older adults with measured characteristics compatible with either activity classification, not to very frail or very active people with little common support.

Research implications

Future studies should measure activity repeatedly before and after time zero with devices capable of distinguishing walking, light activity, MVPA, and sedentary time. More frequent health and outcome assessment would help separate recent recovery or deterioration from activity change and identify when disability begins. Pragmatic trials could then test defined, adaptable activity-support strategies in ADL-independent adults with multimorbidity while prespecifying basic ADL disability, death, mobility, and quality of life. Larger samples and external replication are needed before transporting the present overlap-population estimate to broader clinical or population groups.

Conclusions

Among previously insufficiently active, ADL-independent older adults with multimorbidity, point estimates favored a lower two-year risk of ADL disability or death after attainment of 600 total MET-min/week in the eligible overlap population, although the estimate remained imprecise. The contrast concerns total activity, including walking, and should not be treated as equivalent to an MVPA-only strategy. Replication with larger samples, repeated objective activity measurement, and clearer temporal separation is warranted.

Abbreviations

ADL

activity of daily living

ATO

average treatment effect in the overlap population

BMI

body mass index

CHARLS

China Health and Retirement Longitudinal Study

CI

confidence interval

ESS

effective sample size

IADL

instrumental activity of daily living

IPAQ

International Physical Activity Questionnaire

IPTW

inverse probability of treatment weighting

MET

metabolic equivalent of task

MI

multiple imputation

MICE

multiple imputation by chained equations

MVPA

moderate-to-vigorous physical activity

OW

overlap weighting

PA

physical activity

PS

propensity score

RD

risk difference

RR

risk ratio
SMD
standardized mean difference.

Declarations

Ethics approval and consent to participate

This study was a secondary analysis of registered-access, de-identified data from the 2015, 2018, and 2020 waves of the China Health and Retirement Longitudinal Study (CHARLS). The CHARLS Wave 4 User's Guide states that all CHARLS waves received approval from the Peking University Institutional Review Board. The main household survey, including anthropometric measurements, was approved under IRB00001052-11015, and biomarker collection was approved under IRB00001052-11014. All participants provided written informed consent, with separate consent procedures for the main fieldwork, non-blood biomarkers, blood collection, and blood storage.[18] The present study involved no direct participant contact and used only de-identified secondary data; therefore, additional institutional ethical approval and additional participant consent were not required.

Consent for publication

Not applicable. The manuscript contains no identifiable individual-level information.

Competing interests

The authors declare that they have no competing interests.

Authors' information

Not applicable.

Clinical trial number

Not applicable

Additional files

Additional file 1 (.docx). Supplementary methods and figures.

Contains the multiple-imputation trace and distribution diagnostic (Figure S1) and the temporal structure and sources of identifying uncertainty (Figure S3), with complete legends.

Additional file 2 (.csv). Table S1: Covariate missingness.

Reports missing counts and percentages overall and by 2018 activity strategy.

Additional file 3 (.csv). Table S2: Observed and imputed distributions.

Compares distribution summaries for observed and imputed covariate values.

Additional file 4 (.csv). Table S3: Multiple-imputation diagnostics.

Reports trace, range, constraint, FMI, and Monte Carlo error checks.

Additional file 5 (.csv). Table S4: Propensity-score and overlap-weight diagnostics.

Summarizes score tails, weights, ESS, convergence, and balance across imputations.

Additional file 6 (.csv). Table S5: Known versus unknown outcome comparison.

Describes baseline characteristics for 271 known and 14 unknown outcomes without significance tests.

Additional file 7 (.csv). Table S6: Unknown-outcome extreme scenarios.

Reports four prespecified deterministic assignments and corresponding weighted risks, RRs, and RDs.

Additional file 8 (.csv). Table S7: Full sensitivity-analysis results.

Reports group counts, events, ESS, weighted risks, effect estimates, and changes in exposure, eligibility, weighting, covariates, or outcome handling.

Funding

The authors received no specific funding for this work.

Author Contribution

GL contributed to conceptualization, methodology, data curation, formal analysis, visualization, and preparation of the original draft. QC contributed to conceptualization, methodology, supervision, validation, interpretation of the findings, and review and editing of the manuscript. Both authors read and approved the final manuscript and agree to be accountable for the work.

Acknowledgement

The authors thank the China Health and Retirement Longitudinal Study team and all study participants. The authors also acknowledge the Gateway to Global Aging Data for providing access to Harmonized CHARLS resources used for variable cross-checking and sensitivity analyses.

Data Availability

The de-identified CHARLS data analysed in this study are available to registered users through the official CHARLS data portal (<https://charls.charlsdata.com/pages/data/111/en.html>), subject to application and data-use requirements. Harmonized CHARLS variables used for cross-checking and sensitivity analyses are available

through the Gateway to Global Aging Data (<https://g2aging.org/>) under its registration and use conditions. The authors are not permitted to redistribute individual-level CHARLS records. Analysis code, variable crosswalks, aggregate quality-control outputs, and reproduction instructions are available from the corresponding author on reasonable request, subject to the CHARLS data-use agreement.

References

1. 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(1):61 – 8. doi:10.1093/ije/dys203.
2. Paterson DH, Warburton DE. Physical activity and functional limitations in older adults: a systematic review related to Canada's Physical Activity Guidelines. *Int J Behav Nutr Phys Act*. 2010;7:38. doi:10.1186/1479-5868-7-38.
3. Roberts CE, Phillips LH, Cooper CL, Gray S, Allan JL. Effect of Different Types of Physical Activity on Activities of Daily Living in Older Adults: Systematic Review and Meta-Analysis. *J Aging Phys Act*. 2017;25(4):653–670. doi:10.1123/japa.2016-0201.
4. Delpino FM, de Lima APM, da Silva BGC, Nunes BP, Caputo EL, Bielemann RM. Physical Activity and Multimorbidity Among Community-Dwelling Older Adults: A Systematic Review With Meta-Analysis. *Am J Health Promot*. 2022;36(8):1371–1385. doi:10.1177/08901171221104458.
5. Bricca A, Harris LK, Jäger M, Smith SM, Juhl CB, Skou ST. Benefits and harms of exercise therapy in people with multimorbidity: A systematic review and meta-analysis of randomised controlled trials. *Ageing Res Rev*. 2020;63:101166. doi:10.1016/j.arr.2020.101166.
6. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381-95. doi:10.1249/01.MSS.0000078924.61453.FB.
7. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. *Int J Behav Nutr Phys Act*. 2011;8:115. doi:10.1186/1479-5868-8-115.
8. Sattler MC, Jaunig J, Tösch C, Watson ED, Mokkink LB, Dietz P, et al. Current Evidence of Measurement Properties of Physical Activity Questionnaires for Older Adults: An Updated Systematic Review. *Sports Med*. 2020;50(7):1271–1315. doi:10.1007/s40279-020-01268-x.
9. Zhang L, Chen W, Miao H, Zou T, Xiang X, Wu R, et al. Association between physical activity levels and mild cognitive impairment in Chinese older adults: a cross-sectional study from the China health and retirement longitudinal study. *Front Public Health*. 2025;13:1564544. doi:10.3389/fpubh.2025.1564544.
10. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med*. 2020;54(24):1451–1462. doi:10.1136/bjsports-2020-102955.
11. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758 – 64. doi:10.1093/aje/kwv254.
12. Cashin AG, Hansford HJ, Hernán MA, Swanson SA, Lee H, Jones MD, et al. Transparent reporting of observational studies emulating a target trial: the TARGET Statement. *BMJ*. 2025;390:e087179. doi:10.1136/bmj-2025-087179.

13. Li F, Morgan K, Zaslavsky A. Balancing Covariates via Propensity Score Weighting. *J Am Stat Assoc.* 2018;113(521):390–400. doi:10.1080/01621459.2016.1260466.
14. Li F, Thomas LE, Li F. Addressing Extreme Propensity Scores via the Overlap Weights. *Am J Epidemiol.* 2019;188(1):250–257. doi:10.1093/aje/kwy201.
15. Leyrat C, Seaman SR, White IR, Douglas I, Smeeth L, Kim J, et al. Propensity score analysis with partially observed covariates: How should multiple imputation be used? *Stat Methods Med Res.* 2019;28(1):3–19. doi:10.1177/0962280217713032.
16. Kurland BF, Johnson LL, Egleston BL, Diehr PH. Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims. *Stat Sci.* 2009;24(2):211. doi:10.1214/09-STS293.
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453-7. doi:10.1016/S0140-6736(07)61602-X.
18. China Health and Retirement Longitudinal Study. CHARLS Wave 4 (2018) User's Guide. 2020. https://charls.charlsdata.com/Public/ashelf/public/uploads/document/2018-charls-wave4/application/CHARLS_2018_Users_Guide.pdf. Accessed 23 Aug 2026.
19. China Health and Retirement Longitudinal Study. CHARLS 2015 Household Questionnaire. 2015. https://charls.charlsdata.com/Public/ashelf/public/uploads/document/2015-charls-wave4/application/CHARLS_2015_Questionnaire.pdf. Accessed 23 Aug 2026.
20. Tian Y, Shi Z. Effects of Physical Activity on Daily Physical Function in Chinese Middle-Aged and Older Adults: A Longitudinal Study from CHARLS. *J Clin Med.* 2022;11(21):6514. doi:10.3390/jcm11216514.
21. Sanchez-Sanchez JL, Izquierdo M, Carnicero-Carreño JA, García-García FJ, Rodríguez-Mañas L. Physical activity trajectories, mortality, hospitalization, and disability in the Toledo Study of Healthy Aging. *J Cachexia Sarcopenia Muscle.* 2020;11(4):1007–1017. doi:10.1002/jcsm.12566.
22. Gretebeck RJ, Ferraro KF, Black DR, Holland K, Gretebeck KA. Longitudinal change in physical activity and disability in adults. *Am J Health Behav.* 2012;36(3):385 – 94. doi:10.5993/AJHB.36.3.9.
23. Shah RC, Buchman AS, Leurgans S, Boyle PA, Bennett DA. Association of total daily physical activity with disability in community-dwelling older persons: a prospective cohort study. *BMC Geriatr.* 2012;12:63. doi:10.1186/1471-2318-12-63.
24. Martinez-Gomez D, Bandinelli S, Del-Panta V, Talluri E, Guralnik JM, Ferrucci L. Three-Year Changes in Physical Activity and Subsequent Loss of Ability to Walk 400 m in Older Adults: The InCHIANTI Study. *Am J Phys Med Rehabil.* 2018;97(3):212–215. doi:10.1097/PHM.0000000000000841.
25. Song J, Gilbert AL, Chang RW, Pellegrini CA, Ehrlich-Jones LS, Lee J, et al. Do Inactive Older Adults Who Increase Physical Activity Experience Less Disability: Evidence From the Osteoarthritis Initiative. *J Clin Rheumatol.* 2017;23(1):26–32. doi:10.1097/RHU.0000000000000473.
26. Dunlop DD, Song J, Semanik PA, Sharma L, Bathon JM, Eaton CB, et al. Relation of physical activity time to incident disability in community dwelling adults with or at risk of knee arthritis: prospective cohort study. *BMJ.* 2014;348:g2472. doi:10.1136/bmj.g2472.
27. Mou C, Wang Z, Ke Z. Association between physical activity and multimorbidity: a population-based cohort study. *Arch Public Health.* 2025;83(1):71. doi:10.1186/s13690-025-01562-y.
28. Qiao Y, Liu S, Li G, Lu Y, Wu Y, Shen Y, et al. Longitudinal Follow-Up Studies on the Bidirectional Association between ADL/IADL Disability and Multimorbidity: Results from Two National Sample Cohorts of Middle-

Aged and Elderly Adults. *Gerontology*. 2021;67(5):563–571. doi:10.1159/000513930.

29. Pahor M, Guralnik JM, Ambrosius WT, Blair S, Bonds DE, Church TS, et al. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. *JAMA*. 2014;311(23):2387-96. doi:10.1001/jama.2014.5616.
30. Manini TM, Beavers DP, Pahor M, Guralnik JM, Spring B, Church TS, et al. Effect of Physical Activity on Self-Reported Disability in Older Adults: Results from the LIFE Study. *J Am Geriatr Soc*. 2017;65(5):980–988. doi:10.1111/jgs.14742.
31. Fielding R, Guralnik J, King A, Pahor M, McDermott M, Tudor-Locke C, et al. Dose of physical activity, physical functioning and disability risk in mobility-limited older adults: Results from the LIFE study randomized trial. *PLoS One*. 2017;12(8):e0182155. doi:10.1371/journal.pone.0182155.
32. Penninx BW, Messier SP, Rejeski WJ, Williamson JD, DiBari M, Cavazzini C, et al. Physical exercise and the prevention of disability in activities of daily living in older persons with osteoarthritis. *Arch Intern Med*. 2001;161(19):2309-16. doi:10.1001/archinte.161.19.2309.
33. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act*. 2008;5:56. doi:10.1186/1479-5868-5-56.
34. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol*. 2005;34(1):215 – 20. doi:10.1093/ije/dyh299.
35. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41(3):861 – 70. doi:10.1093/ije/dyr213.
36. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ*. 2007;334(7597):786. doi:10.1136/bmj.39136.682083.AE.
37. Nguyen B, Clare P, Mielke GI, Brown WJ, Ding D. Physical activity across midlife and health-related quality of life in Australian women: A target trial emulation using a longitudinal cohort. *PLoS Med*. 2024;21(5):e1004384. doi:10.1371/journal.pmed.1004384.
38. Martínez-Hernández BM, Rosas-Carrasco O, López-Teros M, González-Rocha A, Muñoz-Aguirre P, Palazuelos-González R, et al. Association between physical activity and physical and functional performance in non-institutionalized Mexican older adults: a cohort study. *BMC Geriatr*. 2022;22(1):388. doi:10.1186/s12877-022-03083-7.
39. Md Sharif S, Jawahir S, Fun WH, Tan EH, Md Razman AH, Awang H, et al. Predictors of transition from independence to limitations in activities of daily living (ADL) among independent community-dwelling older adults: longitudinal evidence from the Malaysia ageing and retirement survey (MARS). *BMC Geriatr*. 2025;25(1):447. doi:10.1186/s12877-025-06076-4.

Figures

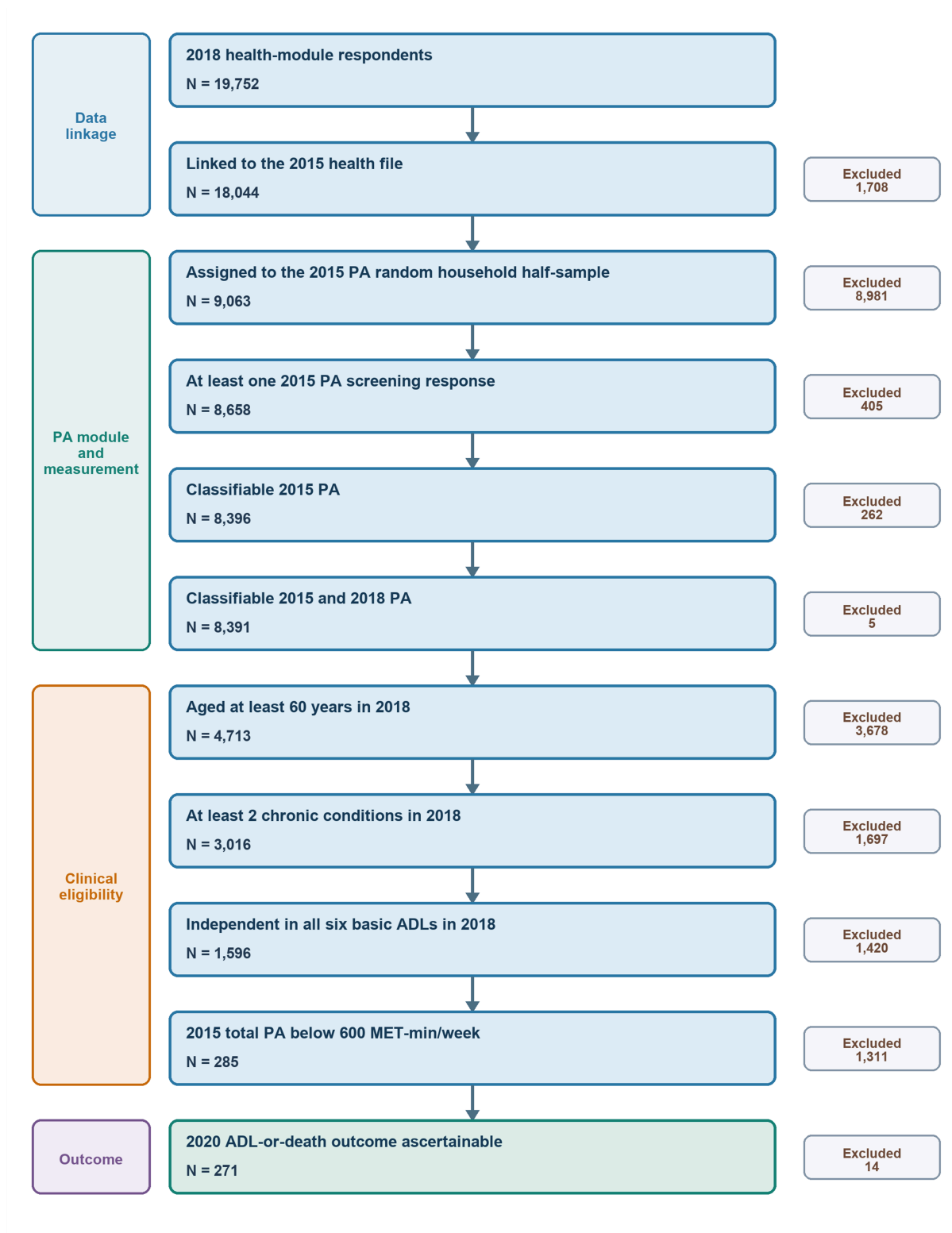


Figure 1

Participant selection flow. Counts are unweighted participants. The household-level 2015 PA random-half module, item response, PA classifiability, clinical eligibility, prior insufficiency, and outcome ascertainment are shown separately. ADL, activity of daily living; PA, physical activity.

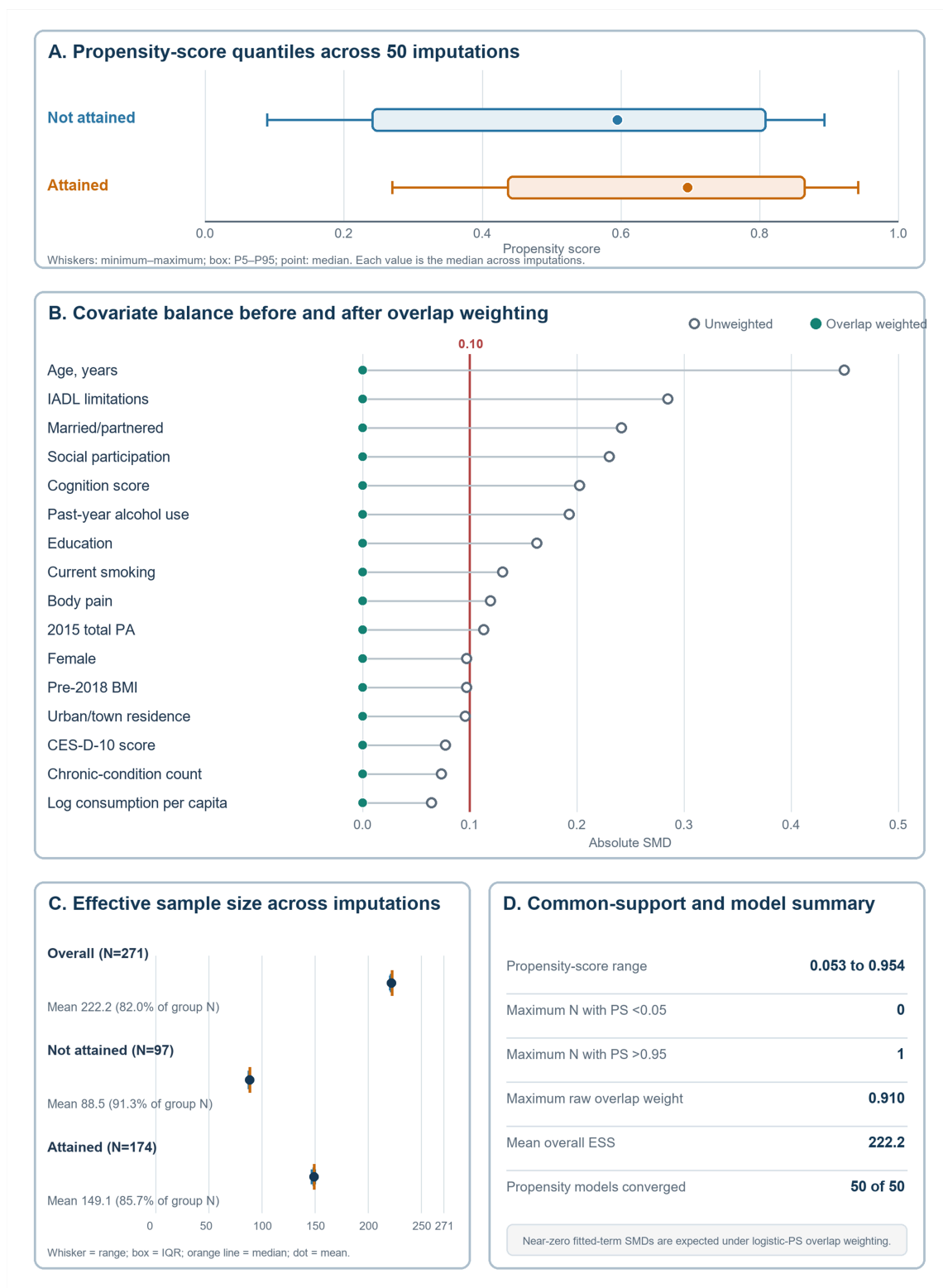


Figure 2

Propensity-score diagnostics across imputations. Panel A shows the median across 50 imputations of the group-specific minimum, P5, median, P95, and maximum propensity scores. Panel B compares mean unweighted and overlap-weighted absolute SMDs. Panel C reports the distribution of overall and strategy-specific ESS; Panel D summarizes common support, maximum OW, and convergence. Near-zero fitted-term

SMDs partly reflect the exact mean-balance property of logistic-PS OW. ESS, effective sample size; MI, multiple imputation; OW, overlap weighting; PS, propensity score; SMD, standardized mean difference.

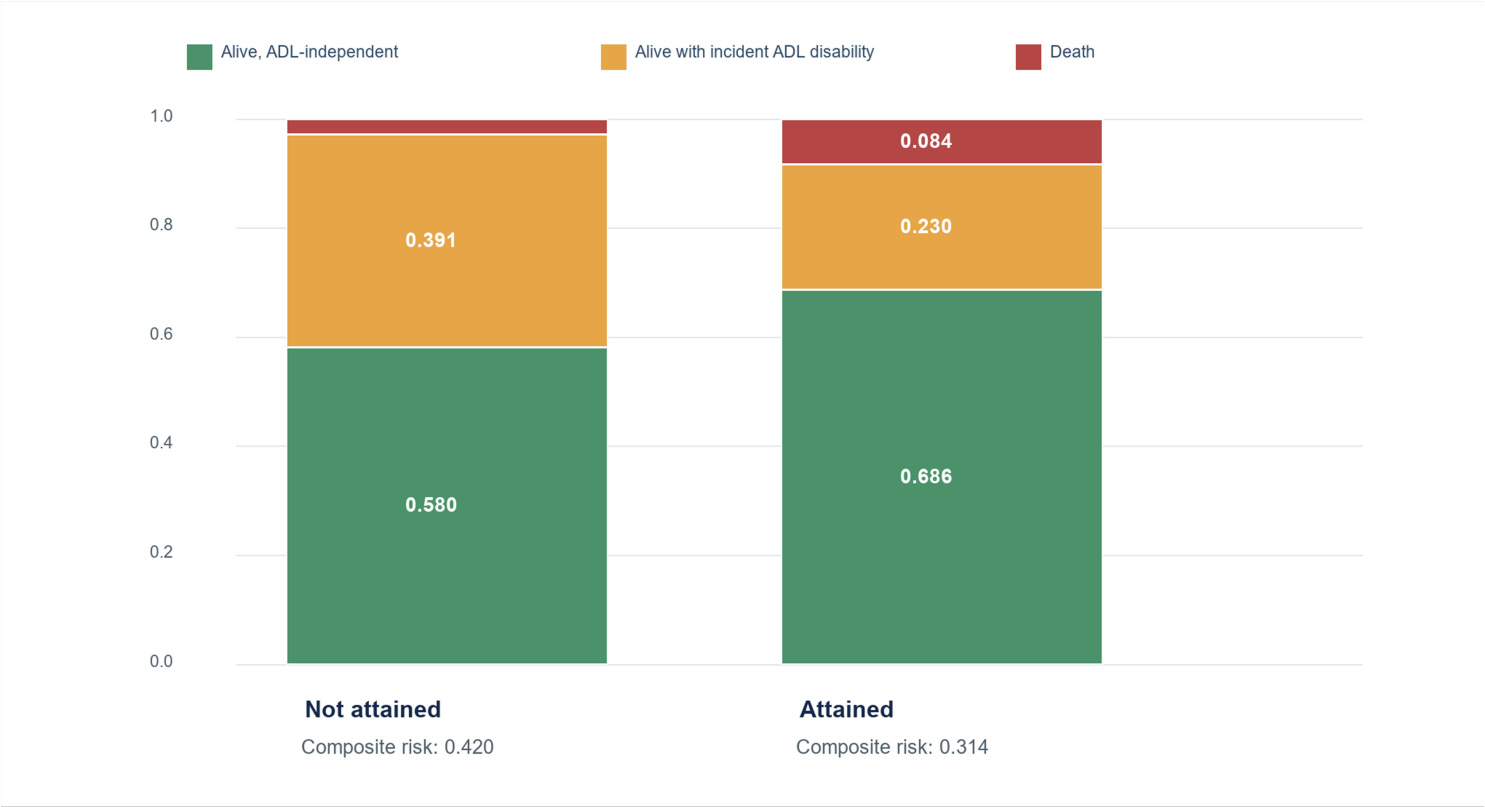


Figure 3

Overlap-weighted composite risks and mutually exclusive outcome-state probabilities. State probabilities use the full baseline denominator and sum to one within strategy. The figure shows how the composite is distributed between alive with incident ADL disability and death. ADL, activity of daily living; OW, overlap weighting.

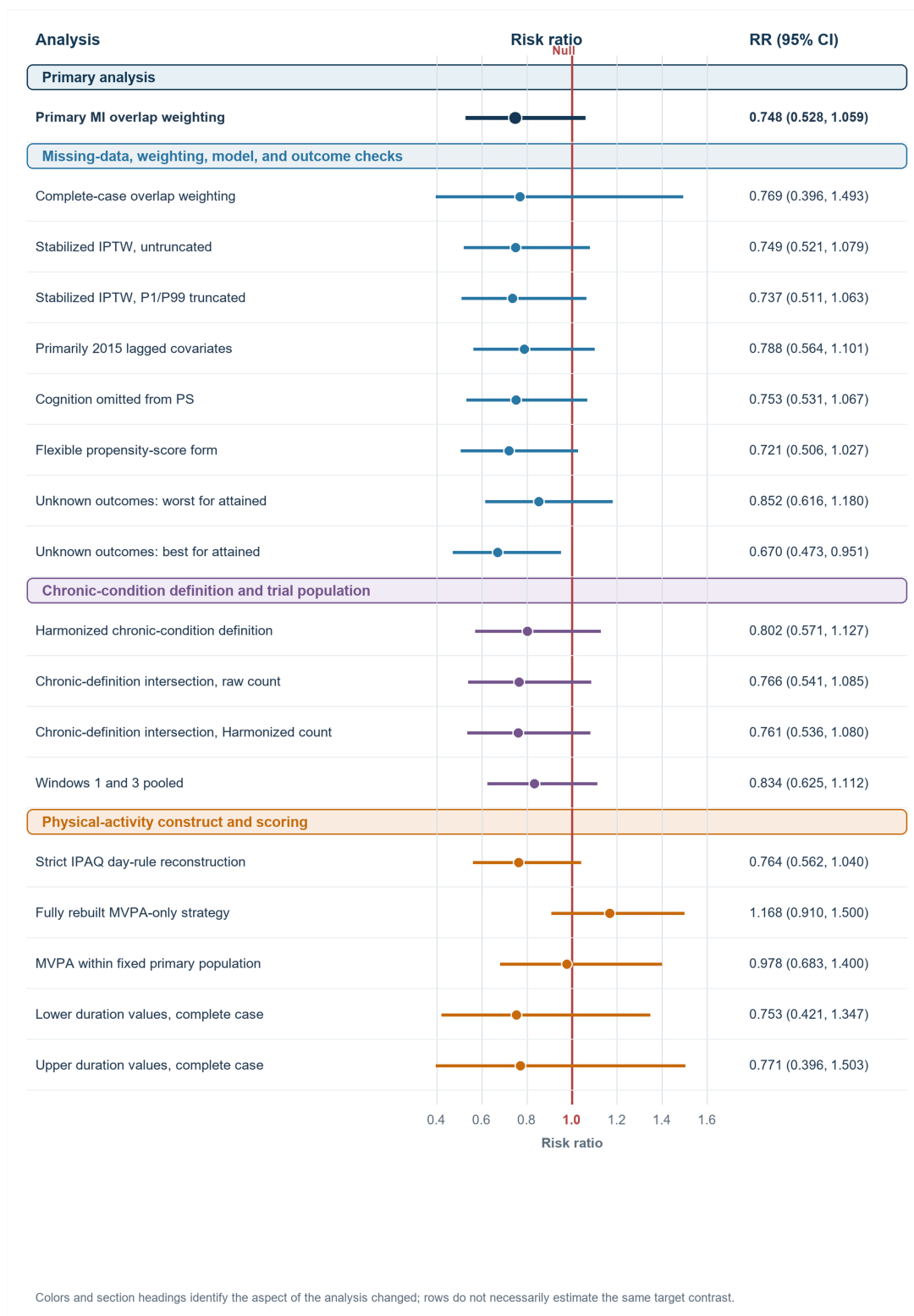


Figure 4

Risk-ratio estimates across prespecified sensitivity analyses. Rows are grouped as the primary analysis; missing-data, weighting, model, and outcome checks; chronic-condition definition and trial-population checks; and physical-activity construct and scoring checks. Different groups need not share one estimand. CI, confidence interval; IPAQ, International Physical Activity Questionnaire; IPTW, inverse probability of treatment weighting; MVPA, moderate-to-vigorous physical activity; OW, overlap weighting; RR, risk ratio.

Supplementary Files

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

- [Additionalfile1supplementarymethodsandfigures.docx](#)
- [Additionalfile2TableS1covariatemissingness.csv](#)
- [Additionalfile3TableS2observedimputeddistributions.csv](#)
- [Additionalfile4TableS3midiagnostics.csv](#)
- [Additionalfile5TableS4PSOWdiagnostics.csv](#)
- [Additionalfile6TableS5knownunknownoutcomes.csv](#)
- [Additionalfile7TableS6unknownoutcomescenarios.csv](#)
- [Additionalfile8TableS7fullsensitivityresults.csv](#)