Supplementary material

Nutritional Indices and Loss of Outdoor Walking after Hip Fracture: A retrospective cohort study of associations and incremental prognostic performance beyond cognition and prefracture mobility

Table S1. Sensitivity analyses

Index	Analysis	n	OR (95% CI)	p
GNRI	Reduced clinical adjustment	212	1.458 (1.029–2.066)	0.034
GNRI	Additional preoperative CRP adjustment	212	1.405 (0.959–2.059)	0.081
GNRI	CRP adjustment excluding postoperative samples (none)	212	1.405 (0.959–2.059)	0.081
GNRI	Excluding older outcome templates	178	1.381 (0.927–2.055)	0.112
GNRI	Exclude 13 chart-confirmed baseline review cases	199	1.471 (1.023–2.117)	0.037
PNI	Reduced clinical adjustment	212	1.323 (1.015–1.725)	0.039
PNI	Additional preoperative CRP adjustment	212	1.277 (0.975–1.673)	0.076
PNI	CRP adjustment excluding postoperative samples (none)	212	1.277 (0.975–1.673)	0.076
PNI	Excluding older outcome templates	178	1.435 (1.033–1.993)	0.031
PNI	Exclude 13 chart-confirmed baseline review cases	199	1.344 (1.032–1.752)	0.028
GNRI	BMI-22 ideal-weight formula	212	1.510 (1.050–2.173)	0.026
PNI	Additional BMI adjustment	212	1.323 (1.018–1.720)	0.037
Albumin	Additionally adjusted for CRP	212	1.517 (1.037–2.219)	0.032
Albumin	Exclude older-template outcomes	178	1.460 (1.013–2.106)	0.043

GNRI effects are per 10-point decrease and PNI effects per five-point decrease. Unless varied, adjustment is identical to the primary models. Reduced adjustment retains age, sex, ASA, cognition and the two walking-aid coefficients. CRP is log(1 + CRP). After review and clinician adjudication, all selected CRP measurements were confirmed to be preoperative, including samples collected on the surgery date. The planned exclusion of postoperative samples therefore removed no patients and duplicates the CRP-adjusted model; it does not provide independent evidence of robustness. The 13 baseline review cases were confirmed to meet eligibility; their exclusion is an additional robustness analysis, not correction of unresolved baseline status. Confidence limits are shown to three decimals to avoid rounding values just above or below 1 to 1.00. The outcome-source sensitivity analysis excluded 34 classifications derived from older structured templates, leaving 178 patients. Albumin effects are per 0.5-g/dL decrease; its CRP and older-template analyses use the same samples and covariates as the corresponding index models.

Nonlinearity likelihood-ratio tests: GNRI p=0.022 and PNI p=0.724. Three-knot restricted cubic splines at the 10th, 50th and 90th percentiles added one nonlinear degree of freedom. The primary GNRI OR is a linear summary of a potentially nonlinear association. Continuous index-by-cognition interactions: GNRI p=0.643 and PNI p=0.275; Holm-adjusted p=0.643 and p=0.549, respectively. Leave-one-out ORs ranged from 1.44 to 1.62 for GNRI and 1.31 to 1.52 for PNI, with p values of 0.011–0.047 and 0.006–0.044, respectively. These analyses do not establish equivalence between cognitive groups. GNRI knots were 76.90, 92.32, and 102.75, and the displayed curve in main Figure 5 uses GNRI 92 as reference. Low-GNRI estimates are imprecise; no range of a constant per-10-point association was validated.

Table S2. Model coefficients for reproducibility

Predictor	Clinical	+ GNRI	+ PNI	+ Albumin
Intercept	-2.039366	-2.328289	-1.995581	1.757414
Age / 10	0.122634	0.136004	0.114635	0.089874
Male sex (0/1)	0.015111	0.003791	-0.030505	-0.027639
ASA ≥3 (0/1)	0.333377	0.250714	0.184216	0.252915
Prefracture cane use (0/1)	1.055581	0.960741	1.018024	0.916806
Prefracture other aid use (0/1)	1.654430	1.479239	1.465731	1.394197
Intertrochanteric/basicervical (0/1)	0.349776	0.315180	0.340999	0.346868
eGFR / 30	-0.454594	-0.489696	-0.476489	-0.478879
Cognitive impairment (0/1)	1.580229	1.431510	1.471428	1.409977
(100 − GNRI) / 10	—	0.409521	—	—
(45 − PNI) / 5	—	—	0.292697	—
−Albumin / 0.5	—	—	—	0.467332

Coefficients are from the apparent unpenalized models. Predicted probability = 1 / [1 + exp(−linear predictor)]. Coding references: female, ASA <3, no walking aid, femoral neck fracture, and no cognitive impairment. Age is scaled per 10 years and eGFR per 30 mL/min/1.73 m²; albumin is in g/dL. These equations have not been externally validated and are not proposed as clinical calculators. All predictors were available in early postoperative care once cognitive status and body measurements were known. ASA was assigned at the anesthetic assessment for surgery. Cognitive impairment denotes an AMTS ≤6 after surgery or a history of dementia in the admission record.

Software: Python 3.12.14, pandas 2.2.3, NumPy 2.3.5, SciPy 1.17.0, statsmodels 0.14.6. Random seed: 20261020. Numerical estimates were checked against binomial generalized linear models; index derivations and cohort totals were checked separately. Revision analyses reproduced the original estimates before adding paired two-stage bootstrap intervals, ridge validation, and additional sensitivity analyses. The revision environment used numpy 2.3.5, pandas 2.2.3, scipy 1.17.0, statsmodels 0.14.6, sklearn 1.8.0. Revision bootstrap seed: 20261119; outer two-stage seed: 20261120.

PNI implementation: absolute lymphocytes (/μL) = recorded WBC × 1000 × lymphocyte percentage / 100. PNI = 10 × albumin (g/dL) + 0.005 × absolute lymphocytes. WBC is expressed in 10^3/μL and Ly in percent, consistent with the institutional laboratory manual. For every patient, albumin, white blood cell count and lymphocyte differential were obtained on the same calendar day. GNRI uses the Lorentz ideal weight with the actual/ideal weight ratio capped at 1; the alternative sensitivity formula uses ideal weight = 22 × height $(m)^2$. Albumin and the laboratory components of PNI were obtained before surgery. The height and weight used for GNRI were generally measured after surgery.

Table S3. Full internal validation metrics

Evaluation	Model	AUC	Brier	Calibration intercept	Calibration slope	CITL
Apparent	Clinical	0.812	0.173	0.000	1.000	0.000
Apparent	+ GNRI	0.821	0.167	0.000	1.000	0.000
Apparent	+ PNI	0.823	0.169	0.000	1.000	0.000
Apparent	+ Albumin	0.827	0.166	0.000	1.000	0.000
Bootstrap corrected	Clinical	0.783	0.190	-0.033	0.857	0.014
Bootstrap corrected	+ GNRI	0.790	0.186	-0.036	0.846	0.010
Bootstrap corrected	+ PNI	0.791	0.188	-0.034	0.837	0.014
Bootstrap corrected	+ Albumin	0.796	0.185	-0.036	0.844	0.009
Repeated CV	Clinical	0.774	0.191	-0.066	0.777	0.008
Repeated CV	+ GNRI	0.780	0.188	-0.072	0.755	0.001
Repeated CV	+ PNI	0.778	0.190	-0.068	0.749	0.010
Repeated CV	+ Albumin	0.784	0.188	-0.074	0.753	-0.001

All models use 212 patients with 87 events. Paired bootstrap optimism correction was updated to 2000 requested resamples; 1991 were successful for all four models. A failed draw would be omitted for all models. Stratified five-fold cross-validation was repeated 20 times with identical folds across models. Metrics were calculated from pooled out-of-fold predictions in each repetition and then averaged; all fits converged. The joint calibration intercept and slope are obtained by regressing outcome on the logit of prediction. Calibration-in-the-large (CITL) instead fixes the slope at 1. Repetition variability is not a confidence interval. Approximate intervals for paired optimism-corrected differences are in Table S6 and do not apply to the repeated-CV estimates. No external evaluation or equivalence test was performed.

Table S4. Laboratory sampling dates

Timing	Laboratory components n (%)	**CRP n (%)**
Relative to admission		
4 days before admission	1 (0.5)	1 (0.5)
1 day before admission	1 (0.5)	0 (0.0)
Admission date	208 (98.1)	210 (99.1)
1 day after admission	1 (0.5)	0 (0.0)
4 days after admission	1 (0.5)	1 (0.5)
Relative to surgery		
Before the surgery date	139 (65.6)	139 (65.6)
On the surgery date, before surgery	73 (34.4)	73 (34.4)
After the surgery date	0 (0.0)	0 (0.0)

Laboratory components of the indices comprise albumin, white blood cell count and lymphocyte differential; all three were selected from the same calendar day for each patient. The denominator is 212 for every column. The admission-date offsets describe the observed selections and were not a prespecified eligibility window. CRP dates were assessed independently. Values and dates reflect review of laboratory records and clinician adjudication of discrepancies, with the original extracted values retained in the audit record. The reviewing clinician confirmed that all samples obtained on the surgery date were collected before the operation; all selected blood samples were therefore preoperative. Percentages may not total 100 because of rounding.

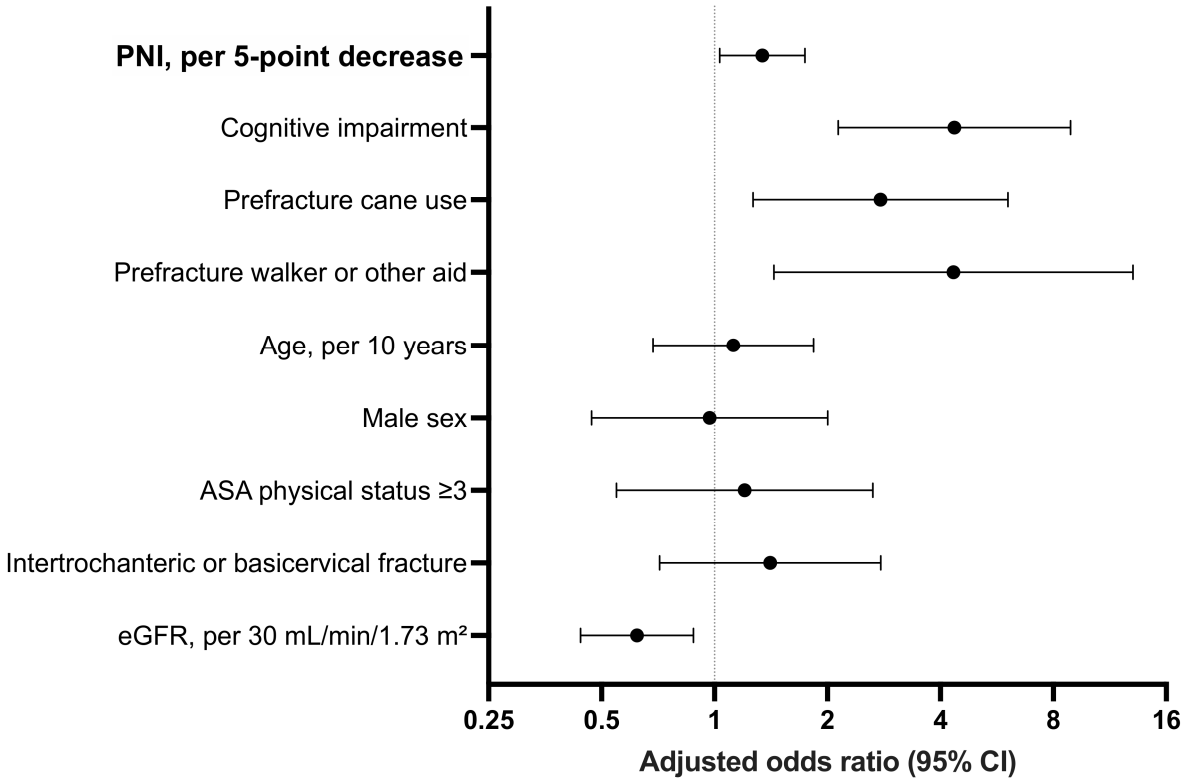

Figure S1. Associations in the PNI model

Points and horizontal bars show adjusted odds ratios and Wald 95% confidence intervals for all nine coefficients in the PNI model (212 patients; 87 losses of outdoor walking). The dotted line indicates an odds ratio of 1. References are no prefracture walking aid, no cognitive impairment, female sex, ASA physical status <3 and femoral neck fracture. PNI is scaled per five-point decrease, age per 10-year increase and eGFR per 30 mL/min/1.73 m² increase. Numerical differences between coefficients from different models are not direct tests of difference. PNI, prognostic nutritional index; ASA, American Society of Anesthesiologists; eGFR, estimated glomerular filtration rate.

Table S5. Early deaths and descriptive outcomes in the full eligible cohort.

Characteristic	120-day survivors (n=212)	Death by 120 days (n=12)
Age, years, median (IQR)	83.0 (78.0–88.0)	89.5 (74.8–94.5)
Women, n (%)	140 (66.0)	8 (66.7)
No prefracture walking aid, n (%)	139 (65.6)	6 (50.0)
Cane, n (%)	48 (22.6)	4 (33.3)
Other walking aid, n (%)	25 (11.8)	2 (16.7)
Intertrochanteric/basicervical fracture, n (%)	83 (39.2)	6 (50.0)
ASA ≥3, n/N available (%)	137/212 (64.6)	9/11 (81.8)
ASA not recorded in study data, n	0	1
Cognitive status in study data, n	212	0

Values are descriptive; no between-group hypothesis tests were performed. Percentages use the group denominator except ASA ≥3, which uses the number with a recorded category. Among survivors, ASA was assigned at the anesthetic assessment for surgery. Two values were confirmed retrospectively during data verification. All 12 patients who died had blood drawn on the day of admission. Their median serum albumin was 3.1 g/dL. Missing cognitive status was not coded as absence of impairment.

Full-cohort outcome	n/N	%
Survived with outdoor walking retained	125/224	55.8
Survived with outdoor walking loss	87/224	38.8
Recorded death within 120 days	12/224	5.4
Death or walking loss (descriptive composite)	99/224	44.2

Three patients were recorded as dead at 30 days; nine were alive at 30 days and recorded as dead at 120 days. These are interval categories derived from recorded survival status, not a distribution of exact death dates. Cognitive status was not abstracted for the 12 patients who died, so we did not fit the composite outcome model with the same covariates as in the primary analysis. The original functional-analysis denominator remains 212.

Table S6. Paired differences in bootstrap optimism-corrected performance.

Comparison	ΔAUC (approx. 95% CI)	ΔBrier (approx. 95% CI)
GNRI vs clinical	0.0079 (-0.0051 to 0.0358)	-0.0040 (-0.0163 to 0.0017)
PNI vs clinical	0.0082 (-0.0053 to 0.0395)	-0.0019 (-0.0170 to 0.0020)
Albumin vs clinical	0.0136 (-0.0037 to 0.0480)	-0.0049 (-0.0208 to 0.0017)
GNRI vs albumin	-0.0057 (-0.0209 to 0.0053)	0.0009 (-0.0030 to 0.0081)
PNI vs albumin	-0.0054 (-0.0269 to 0.0096)	0.0029 (-0.0033 to 0.0116)

Each nutritional model adds the specified marker to the same clinical factors. Differences are model minus reference: positive ΔAUC and negative ΔBrier favor the model named first. Point estimates use paired Harrell optimism correction; intervals use patient-level two-stage percentile bootstrap, repeating model fitting and optimism correction within each outer sample. These are not intervals for cross-validated differences. All models share patient draws; no independence of repeated predictions or resampled rows is assumed.

The point correction used 1991/2000 successful resamples. The interval procedure used 1971/2000 successful outer resamples, each requesting 200 inner resamples; the minimum successful inner count was 162, and the mean was 197.13. Failed inner draws were discarded for every model. Outer draws required successful apparent fits and at least 80% successful inner draws. Point failures: 9; outer failures: 29; failed inner fits across all outer samples: 7538.

These exploratory intervals are pointwise and are not adjusted for multiple comparisons. Approximate percentile intervals have finite Monte Carlo uncertainty and may be affected by small-sample instability, exclusion of non-estimable draws, and nonregular behavior of nested-model comparisons near the null. They do not constitute an equivalence test. Methodological reference: Noma H, Shinozaki T, Iba K, Teramukai S, Furukawa TA. Statistics in Medicine. 2021;40:5691–5701. doi:10.1002/sim.9148.

Table S7. Ridge logistic regression with nested cross-validation.

Marker added	AUC	Brier	Joint intercept	Slope	CITL
clinical	0.774	0.190	0.010	1.022	0.004
GNRI	0.782	0.187	0.006	1.021	0.000
PNI	0.779	0.188	0.014	1.037	0.004
Albumin	0.787	0.185	0.006	1.024	-0.001

The clinical row has no nutritional marker. The outer validation uses the same five-fold partitions repeated 20 times as the unpenalized models. Within each outer training sample, stratified five-fold inner validation selects the inverse penalty C from 25 log-spaced values between 0.001 and 1000 by minimum mean validation log loss. All predictors are standardized within each inner training sample; preprocessing and the selected model are then fitted on the outer training sample. The intercept is not penalized. The outer test sample is used only for evaluation. Reported metrics average the pooled out-of-fold metrics across 20 repetitions. This is a sensitivity analysis of predictive performance, not a replacement for unpenalized association ORs.

Model	Selected C: median (range)	At lower/upper grid boundary
clinical	0.1 (0.03162–0.3162)	0/100; 0/100
GNRI	0.1 (0.03162–0.3162)	0/100; 0/100
PNI	0.1 (0.03162–0.3162)	0/100; 0/100
Alb	0.1 (0.03162–0.3162)	0/100; 0/100

All 100 outer fits per model completed without convergence failure. Joint intercept estimates intercept and slope together; CITL fixes slope at 1. Ridge penalty selection may itself be variable in a small cohort. No external validation or clinical utility analysis is implied.

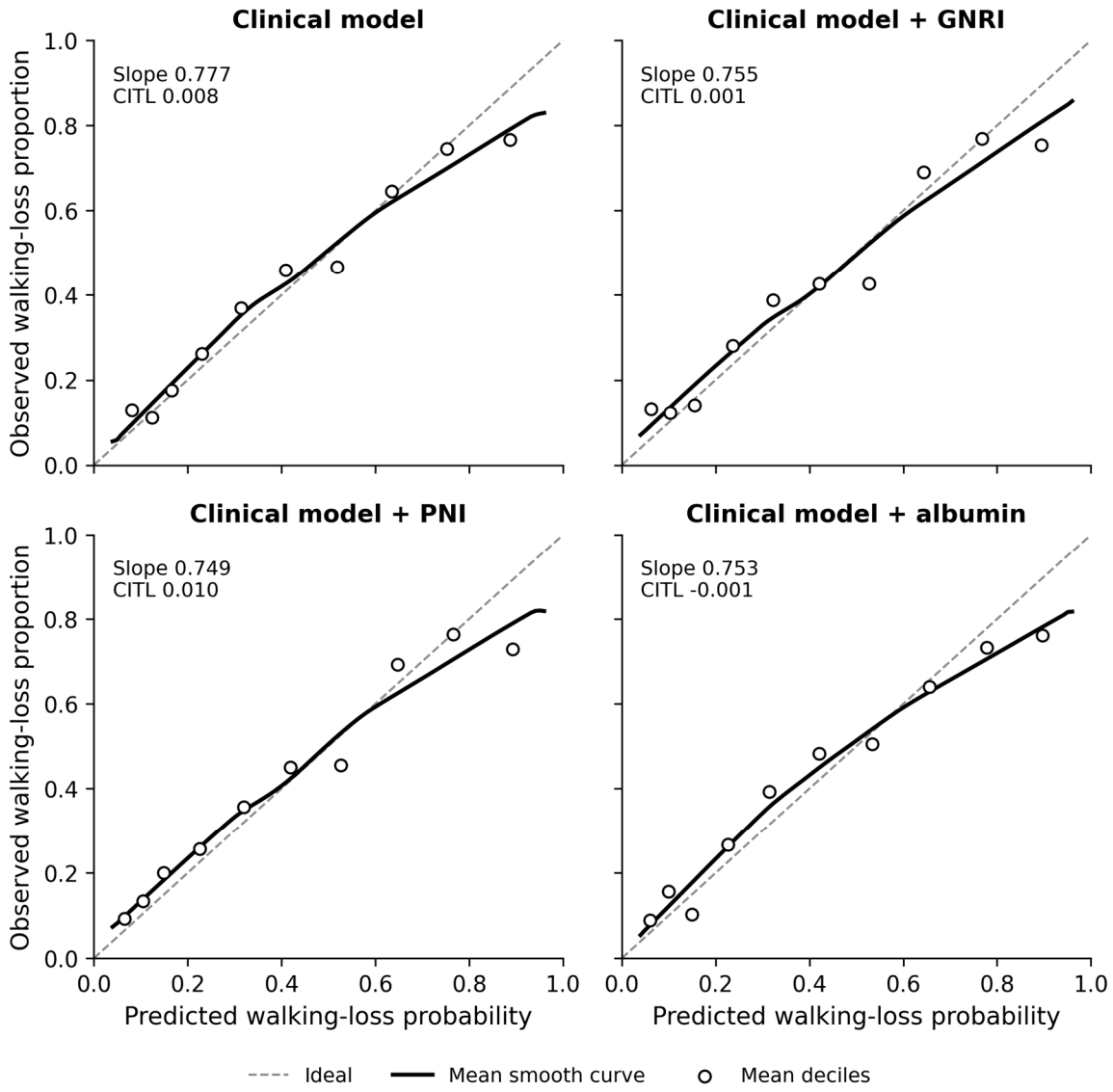

Figure S2. Out-of-fold calibration of unpenalized models.

Each panel uses held-out predictions from five-fold cross-validation repeated 20 times. Solid lines average within-repetition LOWESS curves (span 0.7); open circles average decile summaries across repetitions. The dashed line is ideal calibration. Slope and calibration-in-the-large (CITL) are the means of metrics estimated separately within each repetition. Repetitions are not treated as independent patients. Curves are descriptive and do not show confidence bands.

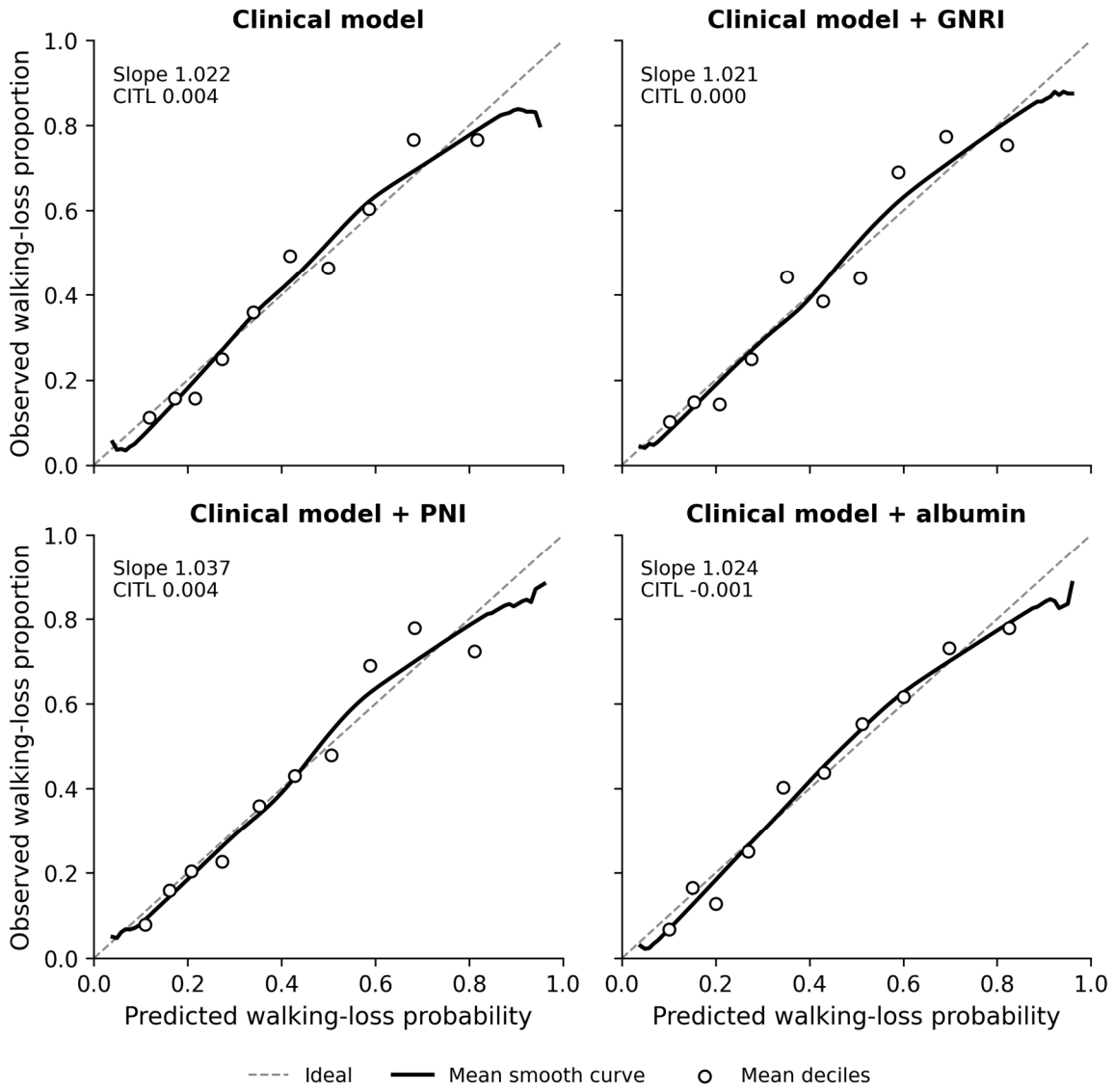

Figure S3. Out-of-fold calibration of ridge models.

Each panel uses held-out predictions from five-fold cross-validation repeated 20 times. Solid lines average within-repetition LOWESS curves (span 0.7); open circles average decile summaries across repetitions. The dashed line is ideal calibration. Slope and calibration-in-the-large (CITL) are the means of metrics estimated separately within each repetition. Repetitions are not treated as independent patients. Curves are descriptive and do not show confidence bands.