Supplementary Material

CT-Derived Skeletal Muscle Index for Prediction of Low Bone Mineral Density

Supplementary Methods. Bootstrap and calibration procedures

Model comparison. Differences in the area under the receiver operating characteristic (ROC) curve (AUC) between nested models were assessed by paired bootstrap. Patients were resampled with replacement (n = 146 per resample, 3,000 resamples); within each resample both models of a nested pair were refitted from scratch so that the comparison was paired, and the difference in AUC was recorded. Ninety-five percent confidence intervals were taken as the 2.5th and 97.5th percentiles of the resulting distribution, and two-sided p-values computed as $2 \times \min(\text{proportion of differences} \leq 0, \text{proportion} \geq 0)$. Confidence intervals for the individual model AUCs were obtained by the same bootstrap percentile procedure. The minimum nonzero increment of the two-sided bootstrap p-value is approximately 2/3,000.

Calibration. The calibration intercept and slope were obtained from the regression of the outcome on the logit of the predicted probability. The Hosmer–Lemeshow test was computed at g = 5, 8, 10 and 12 groups, formed by quantiles of predicted probability with ties broken by rank order. Because apparent calibration measures are optimistically biased when assessed in the development sample, optimism-corrected estimates of the AUC and calibration slope were obtained by Harrell's bootstrap procedure with 1,000 resamples: in each resample the model was refitted, the statistic computed in both the bootstrap sample and the original data, and the mean difference taken as the estimate of optimism.

Software. Analyses were performed in IBM SPSS Statistics v26.0 and in Python (NumPy, SciPy, statsmodels) for bootstrap resampling and calibration analysis. The random seed was fixed for all resampling procedures, and results were independently reproduced from the source dataset.

Supplementary Table S1. Age-restricted sensitivity analysis (patients aged ≥50 years)

All four logistic regression models were refitted in patients aged ≥50 years only (n = 107, 62 events), the age range in which the World Health Organization (WHO) densitometric classification is applicable. Analytical methods were identical to the primary analysis.

Model	Predictor	OR	95% CI	p	AUC
Model 1: Age + Sex	Age	1.076	1.024–1.131	0.004	0.705
	Male sex	0.286	0.100–0.819	0.020	
Model 2: Age + Sex + SMI	Age	1.066	1.011–1.124	0.018	0.777
	Male sex	0.420	0.129–1.369	0.150	
	L3 SMI	0.903	0.846–0.964	0.002	
Model 3: Age + Sex + BMI	Age	1.061	1.007–1.118	0.026	0.787
	Male sex	0.175	0.055–0.560	0.003	
	BMI	0.871	0.795–0.955	0.003	
Model 4: Age + Sex + BMI + SMI	Age	1.057	1.002–1.116	0.043	0.801
	Male sex	0.272	0.075–0.982	0.047	

Model	Predictor	OR	95% CI	p	AUC
	BMI	0.913	0.822–1.014	0.090	
	L3 SMI	0.933	0.867–1.005	0.067	

BMD = bone mineral density; OR = odds ratio; CI = confidence interval; AUC = area under the ROC curve; SMI = skeletal muscle index; BMI = body mass index. L3 = third lumbar vertebra; ΔAUC = difference in AUC. All p values are considered two-sided. Outcome: low BMD (total femur T-score ≤ −1.0). Paired bootstrap comparisons (3,000 resamples): Model 2 vs Model 1 ΔAUC 0.072 (95% CI 0.012–0.150), p = 0.011; Model 4 vs Model 3 ΔAUC 0.014 (95% CI −0.010–0.052), p = 0.35. The incremental discrimination of SMI over age and sex was preserved in this subgroup; the SMI coefficient in the fully adjusted model was no longer nominally significant.

Supplementary Table S2. Threshold sensitivity analysis for the definition of low SMI

Group comparisons were repeated using an alternative threshold set derived from healthy kidney donor candidates rather than from patients with malignancy. Independent samples t-tests throughout, consistent with Table 3.

Primary thresholds (Prado et al.: men <52.4, women <38.5 cm²/m²) — low SMI n = 39 (26.7%)

	Low SMI	Normal SMI	p
Lumbar BMD (g/cm²)	0.96 ± 0.15	1.06 ± 0.17	0.004
Femoral BMD (g/cm²)	0.81 ± 0.18	0.91 ± 0.18	0.005
BMI (kg/m²)	23.95 ± 4.87	29.59 ± 5.78	< 0.001
Postmenopausal women (low SMI n = 15; normal SMI n = 71)			
Lumbar BMD (g/cm²)	0.89 ± 0.12	1.02 ± 0.17	0.003
Femoral BMD (g/cm²)	0.71 ± 0.21	0.88 ± 0.19	0.002

Alternative thresholds (Derstine et al.: men <45.4, women <34.4 cm²/m²) — low SMI n = 21 (14.4%)

	Low SMI	Normal SMI	p
Lumbar BMD (g/cm²)	0.93 ± 0.12	1.05 ± 0.17	0.004
Femoral BMD (g/cm²)	0.79 ± 0.13	0.90 ± 0.19	0.020
BMI (kg/m²)	22.64 ± 4.49	29.00 ± 5.83	< 0.001
Postmenopausal women (low SMI n = 7; normal SMI n = 79)			
Lumbar BMD (g/cm²)	0.88 ± 0.15	1.01 ± 0.17	0.052
Femoral BMD (g/cm²)	0.70 ± 0.12	0.87 ± 0.20	0.036

BMD = bone mineral density; BMI = body mass index; SMI = skeletal muscle index. SD = standard deviation. Values are mean ± SD. All p values are considered two-sided. The direction and significance of the association between low SMI and BMD were preserved in the whole cohort under both threshold sets. In postmenopausal women the femoral difference remained significant under the alternative thresholds

while the lumbar difference became borderline, consistent with the smaller group size (n = 7) rather than with absence of association. The near two-fold difference in estimated prevalence between threshold sets is discussed as evidence of the need for population-specific reference values.

Supplementary Table S3. Calibration of the fully adjusted model (Model 4)

Measure	Value
Brier score	0.181
Apparent AUC	0.803
Optimism-corrected AUC	0.783
Apparent calibration intercept	−0.000 (SE 0.196)
Apparent calibration slope	1.000 (SE 0.192)
Optimism-corrected calibration slope	0.895
Events per variable	20.0
Hosmer–Lemeshow, g = 5	χ^2 0.635, df 3, p = 0.888
Hosmer–Lemeshow, g = 8	χ^2 6.149, df 6, p = 0.407
Hosmer–Lemeshow, g = 10	χ^2 16.127, df 8, p = 0.041
Hosmer–Lemeshow, g = 12	χ^2 11.539, df 10, p = 0.317

AUC = area under the receiver operating characteristic curve; SE = standard error; df = degrees of freedom. Apparent calibration intercept and slope are 0 and 1 by construction when assessed in the development sample; optimism-corrected estimates were therefore obtained by Harrell's bootstrap procedure (1,000 resamples). The Hosmer–Lemeshow test is reported across several group counts because it is sensitive to the number of groups; three of four specifications indicate adequate fit. Interpretation was based primarily on the optimism-corrected calibration slope and the Brier score, which together indicate modest overfitting consistent with 80 events and 4 predictors.

Supplementary Note. Concordance of T-score and Z-score classification in patients aged <50 years

Total femur Z-scores were available for all 146 patients. Among the 39 patients aged <50 years, complete concordance refers specifically to the thresholds of total femur T-score ≤ −2.5 and Z-score ≤ −2.0, rather than the primary low-BMD threshold of T-score ≤ −1.0: all 4 patients with T ≤ −2.5 also had Z ≤ −2.0 (range −2.0 to −2.4), and none of the 14 patients in the −2.5 < T ≤ −1.0 band had Z ≤ −2.0. There were no discordant borderline cases.

A hybrid outcome definition (T ≤ −1.0 at age ≥50 years; Z ≤ −2.0 at age <50 years) was additionally explored. This definition is more conservative in the younger stratum, because the Z-score system has no intermediate category corresponding to osteopenia; outcome prevalence fell from 54.8% to 45.2%. Under this definition age alone became a strong discriminator (Model 1 AUC 0.799, Model 4 AUC 0.828) and the adjusted SMI coefficient was attenuated to non-significance (OR 0.943, 95% CI 0.882–1.009, p = 0.088). This analysis is reported for transparency and is discussed as a limitation; it is not presented as a co-primary analysis, because the two thresholds are not of equivalent stringency.