Open Access Indonesian Journal of Medical Reviews
Volume 6, Issue 4, 2026 | Article ID: 928 | Pages: 311-326 | e-ISSN 2807-6257
OAIJMR

SYSTEMATIC REVIEW AND META-ANALYSIS

Preoperative Sarcopenia and Major Postoperative Complications After Gastrectomy for Gastric Cancer: Does Adding Muscle Function to Muscle Mass Improve Risk Prediction? A Systematic Review and Meta-Analysis

I Putu Oktayana[1*], Gregorius Iwan Budiman Bomba[2], Nugroho Sofian Aji[3]

[1]Faculty of Medicine, Universitas Wijaya Kusuma Surabaya, Surabaya, Indonesia
[2]Department of Surgery, Faculty of Medicine, Universitas Udayana, Denpasar, Indonesia
[3]General Surgery Program, Faculty of Medicine, Universitas Sebelas Maret, Surakarta, Indonesia

ARTICLE INFO

Article history
Received: June 28, 2026
Revised: July 29, 2026
Accepted: August 26, 2026
Available online: August 28, 2026
Published: August 30, 2026

Keywords
Body composition; Clavien-Dindo classification; gastrectomy; gastric cancer; handgrip strength; postoperative complications; sarcopenia; skeletal muscle index

***Corresponding author**
I Putu Oktayana
E-mail: okta72yana@gmail.com

DOI
https://doi.org/10.37275/oaijmr.v6i4.928

ABSTRACT

Background: Consensus bodies define sarcopenia by low muscle mass plus impaired function, yet most gastrectomy studies use computed-tomography muscle mass alone. Whether adding function improves prediction of major postoperative complications remains uncertain.

Objective: To quantify the association between preoperative sarcopenia and major complications after gastrectomy for gastric cancer and compare function-inclusive with imaging-mass-only definitions.

Case presentation: This systematic review and meta-analysis searched PubMed and Europe PMC through 17 June 2026 for adult gastrectomy cohorts reporting Clavien-Dindo grade IIIa or higher complications. Ten cohorts (4,883 patients) entered the primary pool. Sarcopenia increased the odds of major complications (OR 1.71, 95% CI 1.24–2.35; I^2 = 48.0%; prediction interval 0.73–3.99). Function-inclusive and imaging-mass-only definitions produced nearly identical estimates (OR 1.72, 95% CI 0.91–3.23, four cohorts, versus OR 1.71, 95% CI 1.12–2.62, six cohorts; interaction p = 0.997). The estimate remained stable in leave-one-out, cluster-selection, and exact-binomial sensitivity analyses, and no funnel asymmetry was detected.

Conclusion: Preoperative sarcopenia identifies higher-risk patients after gastrectomy. Adding muscle function did not increase the pooled association, but the function-inclusive subgroup was small and incomplete; the absence of a detected difference should not be interpreted as equivalence. Within-cohort evidence favored multidimensional assessment over a single muscle measure.

1. Introduction

Gastric cancer remains one of the largest contributors to the global cancer burden, with approximately 1.1 million new cases and 770,000 deaths estimated worldwide in 2020 and a projected rise towards 1.8 million cases and 1.3 million deaths by 2040[1]. The burden falls disproportionately on East Asia, where China alone accounts for around 37.0% of cases and 39.4% of deaths although age-standardised rates are falling[2], and where organised screening has been followed by measurable declines in incidence[3]. For all but the earliest tumours, in which endoscopic resection is curative and is standard practice in the countries contributing most of the data analysed here, surgical resection with adequate lymphadenectomy remains the principal curative treatment. The operative approach has changed substantially: randomised trials have established that laparoscopic distal gastrectomy achieves five-year survival equivalent to open surgery in locally advanced disease[4,5], and national audit data show that the diffusion of these trials into routine practice was followed by a significant reduction in overall and severe complications[6].

Against that background, the complications that still occur matter more than ever, because they are no longer merely a perioperative inconvenience. Severe morbidity after gastrectomy is associated with markedly worse long-term survival: in one series, three-year and five-year overall survival were 58.1% and 51.6% in patients who sustained a complication against 82.3% and 73.6% in those who did not[7], and among patients undergoing resection for remnant gastric cancer, a Clavien-Dindo grade IIIa or higher event halved five-year overall survival from 65.6% to 32.0%[8]. Part of that effect is mediated through the delay or omission of adjuvant therapy. Identifying, before the operation, the patients in whom such an event is likely is therefore a question with oncological and not only perioperative consequences.

Sarcopenia has become the most intensively studied preoperative risk marker in this setting. It is attractive because it is measurable, potentially modifiable, and biologically plausible as a determinant of the capacity to withstand a major operation. Its predictive value is not confined to gastric surgery: preoperative sarcopenia has been shown to predict complications and non-cancer mortality after oesophagectomy[9], and a respiratory-muscle formulation of the construct stratified complication risk in a stepwise fashion in patients undergoing resection for non-small-cell lung cancer, from 16.4% in those with normal status to 30.0% in those with respiratory sarcopenia[10]. Preoperative frailty, a broader but overlapping construct, independently predicted major postoperative complications in a multicentre prospective cohort of older surgical patients, with an adjusted odds ratio of 2.61[11].

There is, however, a mismatch between how sarcopenia is defined and how it is measured in surgical practice. Every major consensus body treats sarcopenia as a two-component or three-component construct in which low muscle mass must be accompanied by low muscle strength, and in which poor physical performance marks severity. Applying different consensus definitions to the same people produces strikingly different

results. In a Chinese cohort of 4,500 community-dwelling older adults, prevalence ranged from 11.8% to 57.1% depending on which of six criteria was used[12]; across 1,760 Asian older adults it ranged from 4.8% to 16.1% across nine definitions[13]; and applying four diagnostic frameworks to the same 330 patients with type 2 diabetes gave a prevalence range of 7.58% to 27.27%[14]. Even changing only the functional component matters: within a single framework, substituting a chair-stand test for handgrip strength raised the prevalence of probable sarcopenia from 7.1% to 31.7%, with poor agreement between the two strength criteria[15]. In patients with oesophagogastric cancer, the individual components diverge further still, with low skeletal muscle index and myosteatosis each present in half of patients preoperatively while fully defined sarcopenia affected only 15%[16].

Surgical cohorts have nonetheless converged, for entirely practical reasons, on a single cross-sectional muscle measurement made on a staging computed tomogram that has already been acquired. This is cheap, retrospective and reproducible, but it captures only one of the two constructs the definition requires. Prevalence estimates obtained this way are correspondingly high and variable — 62% in one prospective Indian series[17] and 51.6% in an eighteen-year Finnish cohort[18] — which is what one expects of a measurement whose threshold is inherited from a reference population that may not resemble the patients in front of the surgeon.

The consequence is that the accumulated evidence on sarcopenia and gastrectomy answers a narrower question than the one it appears to answer. The most recent and largest synthesis in this field pooled 42 studies and 11,981 patients, and restricted itself by design to sarcopenia defined by the skeletal muscle mass index[19]. An earlier synthesis of computed-tomography-derived low muscle mass reported an odds ratio of 1.73 for severe complications and closed by calling explicitly for a universal cut-off value, acknowledging that heterogeneity of definition was limiting the field[20]. The one synthesis that did organise itself around consensus-defined sarcopenia examined elective abdominal surgery in general rather than gastrectomy specifically, and reported a substantially larger association in its gastric subgroup[21]. No review has yet compared the two families of definition, head to head, within the gastrectomy population, and no review has examined whether the apparent abundance of function-inclusive evidence in this field is as independent as it appears.

The novelty of this study lies in being the first synthesis in gastric cancer surgery to treat the sarcopenia definition itself as the exposure under investigation rather than as a nuisance source of heterogeneity: it compares, within a single pooled analysis and with a formal test for subgroup difference, definitions that incorporate muscle strength or physical performance against definitions based on imaging muscle mass alone; it supplements that between-study comparison with within-cohort evidence in which two or more definitions were applied to the same patients, so that between-cohort confounding cannot explain any difference; and it is the first to resolve the serially published reports that dominate this literature by institution and accrual window, and to show what that structure does to previously published pooled estimates. The aim of this study was to quantify the association between preoperative sarcopenia and major postoperative complications, defined as Clavien-Dindo grade IIIa or above, in adults undergoing gastrectomy for gastric cancer, and to determine whether that association is stronger when the sarcopenia definition incorporates muscle function than when it rests on imaging-derived muscle mass alone.

2. Methods

2.1. Design and reporting

This systematic review and meta-analysis was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 statement[22]. The review was not prospectively registered, and this is stated here rather than omitted. A protocol specifying the population, exposure, comparator, outcome, primary subgroup and analysis set was fixed before any pooling was undertaken, and every analysis reported below was pre-specified except where it is explicitly labelled as exploratory.

2.2. Search strategy

PubMed and Europe PMC were last searched on 17 June 2026, with a supplementary check of the reference lists of previous syntheses in this field and of every retrieved full text. The core Boolean string executed in PubMed was: (sarcopenia[tiab] OR "skeletal muscle index"[tiab] OR "low muscle mass"[tiab] OR "muscle mass"[tiab] OR "muscle depletion"[tiab] OR myopenia[tiab]) AND (gastrectomy[tiab] OR "gastric cancer"[tiab] OR "stomach neoplasms"[MeSH]) AND (complication[tiab] OR morbidity[tiab] OR Clavien[tiab] OR "surgical outcome"[tiab]). Six variant queries were executed to broaden sensitivity around the exposure and outcome terms, together with an open-access sweep of Europe PMC. Records were restricted to those published in English. Screening was performed by a single reviewer; this is a genuine methodological limitation of the review and is recorded as such in the study-selection diagram rather than presented as a dual-reviewer process.

2.3. Eligibility criteria

Studies were eligible if they enrolled adults undergoing gastrectomy with curative intent for histologically confirmed gastric adenocarcinoma; assessed sarcopenia or low muscle mass before the index operation using a stated instrument and a stated threshold; included a non-sarcopenic comparator group drawn from the same cohort; and reported postoperative complications graded with the Clavien-Dindo classification at a threshold of grade IIIa or above, or provided an odds ratio with a confidence interval for that endpoint. Eligible designs were prospective and retrospective cohort studies and prospective studies nested within randomised trials. Reviews, meta-analyses, editorials, conference abstracts, case reports and case series without a comparator were excluded, as were studies of bariatric surgery, of non-gastric primary tumours, and of postoperative rather than preoperative muscle change. Studies whose exposure was sarcopenic obesity or muscle quality alone were not admitted to the primary pool but were retained for pre-specified secondary analyses.

2.4. Handling of overlapping cohorts

This field contains an unusual number of serially published reports drawn from the same surgical departments, and treating them as independent would inflate precision and could dominate a subgroup. Every eligible report was therefore assigned to a cluster, and one report was admitted to the primary pool from each cluster. A cluster was defined by the combination of institution, accrual window and exposure construct. The third element of that definition requires justification, because it is the point at which this rule departs from a simple institutional rule. What is being pooled in this review is the performance of a definition, not the experience of a patient group; where one department has published two reports describing the same patients but characterising muscle in two genuinely different ways, one by imaging mass and one by measured strength, those reports contribute different exposures to different subgroups and excluding one of them would remove the very contrast the review exists to examine. Both were therefore admitted, and a pre-specified sensitivity analysis collapsing such a cluster to its single largest report was run to demonstrate that the decision does not drive the result. Where two reports from one institution described genuinely non-overlapping accrual periods, both were admitted on that ground alone. Preference within a cluster was otherwise given to the report with the largest sample, the most explicitly stated definition, and directly extractable event counts, and the alternates were retained for a pre-specified cluster-swap analysis.

2.5. Data extraction

For each study the following were extracted: country, institution, design, accrual window, sample size, the sarcopenia instrument

and its numerical threshold exactly as applied, which of the three sarcopenia constructs the definition required, the surgical approach, the Clavien-Dindo threshold used, the number of major complications in each exposure group, and any study-level multivariable estimate together with the variables it adjusted for. Where a two-by-two table was reconstructable, it was used directly; where only an odds ratio with a confidence interval was published, the log odds ratio and its standard error were recovered from the interval. No value was imputed, back-calculated from a p-value alone, or carried forward unverified. Reports whose full text could not be obtained, and from whose abstract the counts could not be recovered, were named in the study-selection diagram rather than estimated.

2.6. Risk of bias and certainty of evidence

Risk of bias was assessed with ROBINS-E, the instrument designed for non-randomised studies of exposures, across its seven domains: confounding, measurement of the exposure, selection of participants, post-exposure interventions, missing data, measurement of the outcome, and selection of the reported result[23]. An instrument for randomised trials would not have been applicable, since every eligible study in this field is an observational cohort. The Newcastle-Ottawa Scale was scored in parallel so that this appraisal remains comparable with the earlier syntheses in this field, which used it; the two instruments measure different things on different scales, are reported separately, and are not interchangeable. Certainty of the evidence was rated using the GRADE approach adapted for prognostic-factor questions, starting at low certainty because the evidence base is entirely observational.

2.7. Statistical analysis

The primary outcome is dichotomous, since a patient either does or does not sustain a Clavien-Dindo grade IIIa complication, and the odds ratio was therefore the effect measure throughout. It is also the measure in which the comparator literature is reported, which permits direct triangulation. Estimates were pooled on the log scale under an inverse-variance random-effects model with the DerSimonian-Laird estimator of between-study variance. A continuity correction of 0.5 was pre-specified for zero cells, although no included study contained one.

Heterogeneity was quantified with τ^2, with I^2 and its confidence interval, with Cochran's Q, and with a 95% prediction interval. Because the function-inclusive subgroup was expected to be small, the Hartung-Knapp adjustment and a restricted-maximum-likelihood estimator of τ^2 were run alongside the primary model, and the Hartung-Knapp interval is quoted wherever the fragility of an estimate is under discussion. Two studies contributed very few events, so the eight studies with reconstructable counts were additionally re-fitted with an exact within-study binomial likelihood using a one-stage binomial-normal generalised linear mixed model, and with the Peto and Mantel-Haenszel fixed-effect estimators, to confirm that the normal approximation to the log odds was not driving the result.

The pre-specified primary subgroup analysis compared function-inclusive against imaging-mass-only definitions with a formal test for subgroup difference. Pre-specified secondary subgroups were geographical region, study design and surgical approach, together with a finer stratification by which sarcopenia constructs each definition actually required. Because five interaction tests were performed, raw interaction p values were corrected using both the Holm and the Benjamini-Hochberg procedures, and no subgroup finding is reported as significant unless it survives correction. Random-effects meta-regression was used to test definition family, publication year and sample size as moderators, and to test a graded moderator counting the number of sarcopenia constructs each definition required. Small-study effects were examined with Egger's regression, the Begg-Mazumdar rank correlation and trim-and-fill, all pre-specified to run only if at least ten studies contributed. Sensitivity analyses comprised leave-one-out removal overall and within each subgroup, the cluster swap described above, restriction to institutionally independent cohorts, restriction to studies with reconstructable counts, and the addition of the sarcopenic-obesity cohorts. A parallel pool of study-level adjusted estimates was pre-specified to assess whether measured confounding could account for the association. Analyses were performed in R.

3. Results

3.1. Study selection

The searches returned 976 records, of which 614 were duplicates or had already been retrieved. Of the 362 records screened on title and abstract, 332 were excluded: 96 were reviews, meta-analyses, editorials or case reports; 121 did not concern gastrectomy for gastric cancer; 64 contained no sarcopenia or muscle-composition exposure; and 51 reported no postoperative complication outcome. Full text was sought for 30 reports and could not be obtained for nine, in every case because the article sat behind a subscription paywall and the event counts could not be recovered from the abstract. Of the 21 reports assessed for eligibility, six were excluded from the primary pool: two because another report from the same institution and accrual window had already been included, two because the exposure was sarcopenic obesity rather than sarcopenia alone, one because the exposure was muscle quality alone, and one because body composition was analysed only as a continuous per-standard-deviation score with no dichotomy available. Fifteen studies entered the review, of which ten contributed to the primary quantitative synthesis and five to sensitivity or secondary analyses. The flow of records from identification to inclusion, including the reason for each exclusion, is detailed in Figure 1.

3.2. Characteristics of the included studies

The ten cohorts in the primary synthesis comprised 4,883 patients operated on in eight institutions across five countries: China, Japan, Turkey, Poland and Finland, with Japanese and Chinese cohorts contributing the largest samples. Two were prospective and eight retrospective. Four applied a definition that incorporated muscle function and six relied on imaging muscle mass alone. Among the function-inclusive cohorts, two applied a full three-component consensus algorithm including gait speed, one applied a handgrip-strength criterion after propensity matching on body composition, and one applied a composite consensus criterion requiring low grip strength together with low muscle mass or low muscle-specific strength, the last of these being the ratio of grip strength to muscle area. Among the imaging-only cohorts, all six used the skeletal muscle index measured at the third lumbar vertebra, but they applied four different threshold sets. Two institutions each contributed two reports to the primary pool: the department reported by Zhuang and Wu and colleagues contributed cohorts accrued in two non-overlapping periods, and the department reported by Matsui and colleagues contributed the same patients characterised once by muscle mass and once by muscle strength, which under the clustering rule set out in Section 2.4 enter different subgroups. Study characteristics, definitions and event counts are set out in Table 1.

Open Access Indonesian Journal of Medical Reviews
Volume 6, Issue 4, 2026 | Article ID: 928 | Pages: 311-326 | e-ISSN 2807-6257
OAIJMR

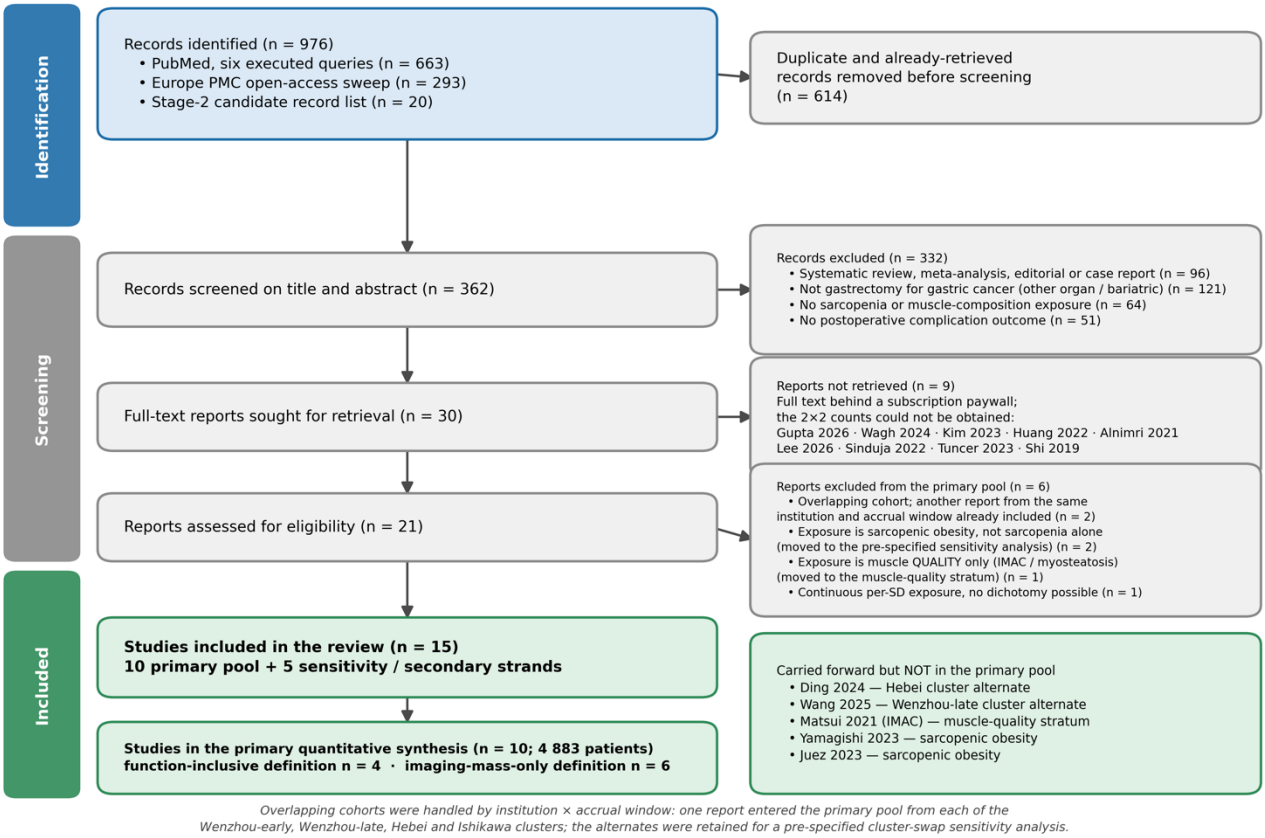

Figure 1. Study selection, presented according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 flow diagram. Overlapping cohorts were resolved by institution and accrual window, so that one report entered the primary pool from each of four clusters; the alternates were retained for a pre-specified cluster-swap sensitivity analysis.

Table 1. Characteristics of the ten cohorts contributing to the primary quantitative synthesis, ordered by definition family and year.

Study	Country	Design	Definition family	Instrument and threshold as applied	N	Events / n (sarcopenic vs not)	Crude OR (95% CI)
Fukuda 2016[24]	Japan	Retrospective	Function-inclusive	EWGSOP algorithm: bioelectrical impedance muscle mass, handgrip strength and gait speed	99	6/21 vs 7/78	4.06 (1.19-13.80)
Erkul 2022[25]	Turkey	Prospective	Function-inclusive	EWGSOP2, all three components measured preoperatively	146	1/31 vs 6/115	0.61 (0.07-5.23)
Matsui 2022[26]	Japan	Retrospective	Function-inclusive	AWGS 2019 low handgrip strength; 1:1 propensity matching on body composition	114	1/57 vs 2/57	0.49 (0.04-5.57)
Wu 2025[27]	China	Prospective	Function-inclusive	GLIS criterion 3: low grip strength plus low skeletal muscle index or low muscle-specific strength	1,654	Two-by-two table not published	1.68 (1.11-2.53)
Zhuang 2016[28]	China	Retrospective	Imaging-mass only	CT L3 skeletal muscle index; optimum-stratification cut-offs 40.8 / 34.9 cm^2/m^2	937	40/389 vs 21/548	2.88 (1.67-4.96)
Sierzega 2019[29]	Poland	Retrospective	Imaging-mass only	CT L3 skeletal muscle index; Prado / Martin international consensus cut-offs	138	22/60 vs 16/78	2.24 (1.05-4.80)
Matsui 2021[30]	Japan	Retrospective	Imaging-mass only	CT L3 skeletal muscle index; Zhuang cut-offs 40.8 / 34.9 cm^2/m^2	840	Two-by-two table not published	0.87 (0.50-1.52)
Back 2025[18]	Finland	Retrospective	Imaging-mass only	CT L3 skeletal muscle index; ≤ 55 cm^2/m^2 men, ≤ 39 cm^2/m^2 women	337	41/174 vs 35/163	1.13 (0.68-1.88)
Guo 2025[31]	China	Retrospective	Imaging-mass only	CT L3 skeletal muscle index; Zhuang cut-offs 40.8 / 34.9 cm^2/m^2	381	22/178 vs 10/203	2.72 (1.25-5.92)
Akmercan 2025[32]	Turkey	Retrospective	Imaging-mass only	CT L3 skeletal muscle index; published sex-specific cut-offs	237	17/87 vs 18/150	1.78 (0.86-3.67)

AWGS, Asian Working Group for Sarcopenia; CT, computed tomography; EWGSOP, European Working Group on Sarcopenia in Older People; GLIS, Global Leadership Initiative in Sarcopenia; L3, third lumbar vertebra; OR, odds ratio. The outcome is a major postoperative complication, Clavien-Dindo grade IIIa or above. Two studies did not publish the underlying two-by-two table but did report an odds ratio with a confidence interval, from which the log-odds and its standard error were recovered; all ten therefore contribute an odds ratio. Observation windows were not uniform: the source reports variously specified 30 days, the index admission, or both, and they did not report enough detail for these to be separated, so a complication arising during a prolonged admission beyond day 30 would have been captured by some cohorts and not by others.

The heterogeneity of thresholds within the imaging-only group, visible in the instrument column of Table 1, deserves emphasis, because it is the practical form the definition problem takes in this literature. Three cohorts applied the cut-offs of 40.8 and 34.9 cm^2/m^2 derived by optimum stratification in a Chinese population[28,30,31], one applied the widely used international consensus values[29], one applied thresholds of 55 and 39 cm^2/m^2[18], and one applied a further set of published sex-specific values[32]. These sets are not interchangeable, and a patient classified as sarcopenic under one may be classified as normal under another.

3.3. Risk of bias

Eight of the ten cohorts were rated as raising some concerns overall and two as being at high risk of bias. Matsui 2022 was rated high risk because handgrip strength had been recorded in only a subset of the source cohort and the analysis rests on a 114-patient propensity-matched subsample, so that the reported result is conditional on a selection step that cannot be fully characterised[26]. Akmercan 2025 was rated high risk because no multivariable model was fitted for the complication endpoint, leaving confounding by age, stage and comorbidity entirely unaddressed[32]. At domain level, confounding was the weakest area, with only 30% of cohorts rated low risk, followed by selection of participants and missing data at 50% each. This pattern is exactly what one expects of a body of retrospective cohorts assembled from patients who happened to have an interpretable preoperative computed tomogram, since those without one are excluded in a manner that is unlikely to be random. Measurement of the outcome was the strongest domain, at 90% low risk, reflecting the consistency with which the Clavien-Dindo classification is applied in this literature. Newcastle-Ottawa scores ranged from six to nine stars. The judgement for every cohort in each of the seven domains, and the summary across studies, is detailed in Figure 2.

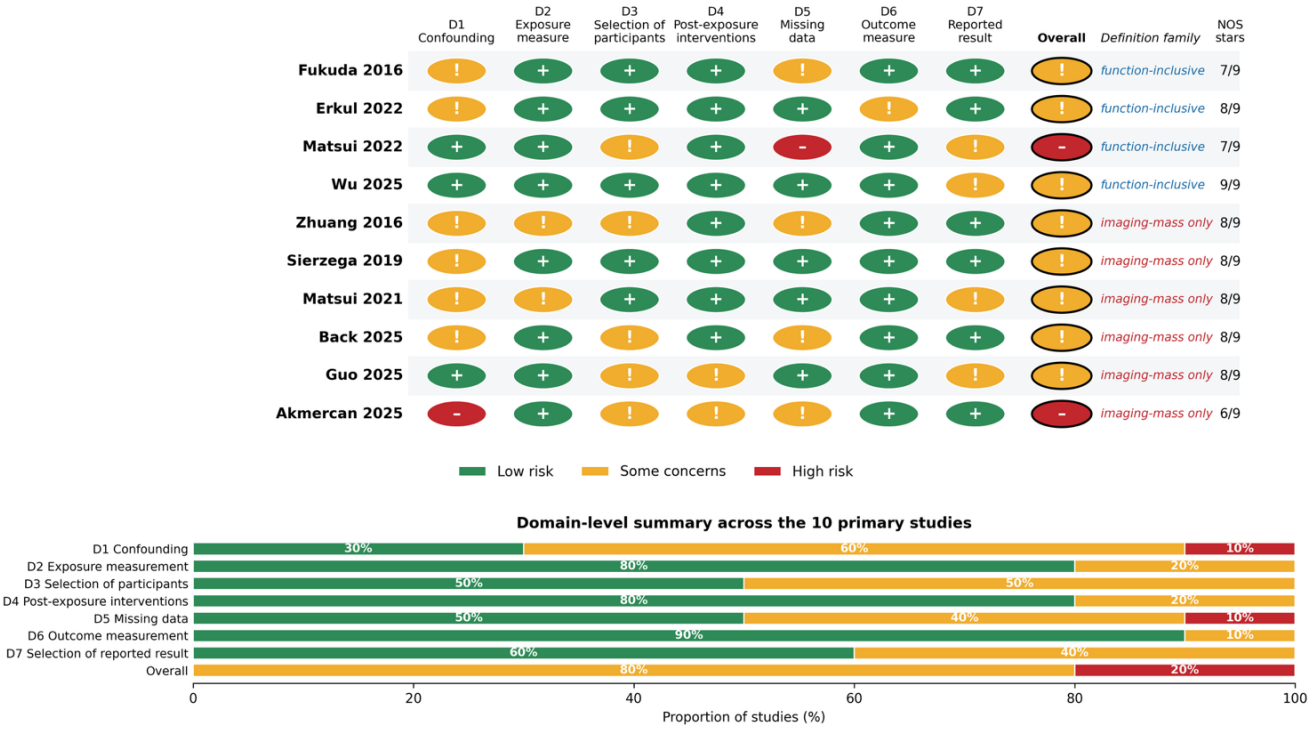

Figure 2. Risk of bias across the seven ROBINS-E domains for the ten cohorts in the primary synthesis, with the domain-level summary beneath. Green denotes low risk, amber some concerns and red high risk. Newcastle-Ottawa scores are shown alongside solely for comparability with earlier syntheses in this field, which used that instrument; the two instruments assess different constructs on different scales and a cohort may be rated favourably on one and unfavourably on the other.

3.4. Association between preoperative sarcopenia and major complications

Preoperative sarcopenia, however defined, was associated with higher odds of a major postoperative complication. The pooled odds ratio was 1.71 (95% confidence interval 1.24 to 2.35; p = 0.0012) across ten cohorts and 4,883 patients. Between-study heterogeneity was moderate and below the pre-specified threshold of 75% that would have mandated a deeper stratified analysis: I^2 was 48.0% (95% confidence interval 0.0% to 74.9%), τ^2 was 0.1138, and Cochran's Q was 17.32 on nine degrees of freedom (p = 0.044). The 95% prediction interval, 0.73 to 3.99, crosses unity, which means that in a new and as yet unstudied setting the direction of the association is not guaranteed. Under the Hartung-Knapp adjustment the point estimate was unchanged but the interval widened to 1.17 to 2.48 (p = 0.010), as the second row of Table 2 shows. The study-level estimates, their weights and the pooled result are detailed in Figure 3, and the pooled estimate under four model specifications is given in Table 2.

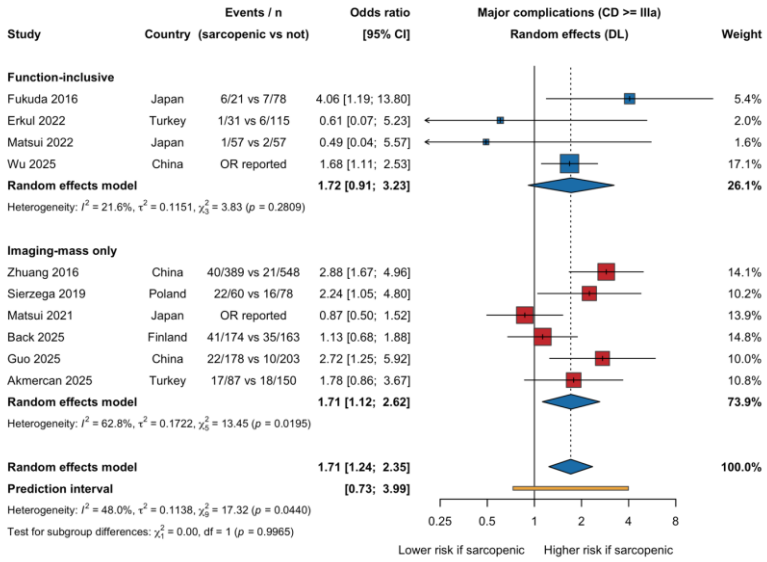

Figure 3. Forest plot of the association between preoperative sarcopenia and major postoperative complications (Clavien-Dindo grade IIIa or above) under an inverse-variance random-effects model with the DerSimonian-Laird estimator, stratified by definition family. Blue denotes function-inclusive definitions and red imaging-mass-only definitions. The orange bar beneath the summary diamond is the 95% prediction interval.

Table 2. The pooled association between preoperative sarcopenia and major postoperative complications under four model specifications.

Model specification	k	Pooled OR	95% CI	p	τ^2	I^2 (%)	95% prediction interval
Random effects, DerSimonian-Laird (primary)	10	1.71	1.24 to 2.35	0.0012	0.1138	48.0	0.73 to 3.99
Random effects, DerSimonian-Laird with Hartung-Knapp	10	1.71	1.17 to 2.48	0.010	0.1138	48.0	—
Random effects, restricted maximum likelihood	10	1.71	1.23 to 2.36	0.0013	0.1174	48.0	—
Common (fixed) effect	10	1.66	1.35 to 2.05	< 0.001	—	48.0	—

CI, confidence interval; k, number of contributing studies; OR, odds ratio. The Hartung-Knapp interval is the more conservative and is the one quoted whenever the fragility of the estimate is discussed.

3.5. The primary comparison: does adding muscle function help?

The pre-specified test for subgroup difference answered the question posed in the title, and the answer was negative. Definitions that incorporated muscle function gave a pooled odds ratio of 1.72 (0.91 to 3.23) across four cohorts, and definitions based on imaging muscle mass alone gave 1.71 (1.12 to 2.62) across six. The two estimates are almost numerically identical and the test for subgroup difference was as close to null as such a test can be, with Q between groups of 0.000 on one degree of freedom (p = 0.997). Random-effects meta-regression agreed: definition family explained none of the between-study variance, with an odds ratio per family of 1.01 (0.45 to 2.27; p = 0.989) and an R^2 of 0%, and it remained null after adjustment for publication year and sample size. A graded moderator counting the number of sarcopenia constructs each definition required, from one to three, was also null, with an odds ratio of 1.14 per additional construct (0.65 to 1.99; p = 0.651). The subgroup analyses are summarised in Table 3.

Table 3. Pre-specified subgroup analyses with tests for subgroup difference, before and after correction for the five comparisons performed.

Subgroup analysis	Level	k	Pooled OR	95% CI	I^2 (%)	p for interaction	p after Holm correction
Definition family (pre-specified primary)	Function-inclusive	4	1.72	0.91 to 3.23	21.6	0.997	1.000
	Imaging-mass only	6	1.71	1.12 to 2.62	62.8		
Geographical region	Asian	6	1.87	1.15 to 3.03	62.6	0.429	1.000
	Western	4	1.46	1.00 to 2.12	4.3		
Study design	Retrospective	8	1.77	1.18 to 2.66	57.5	0.750	1.000
	Prospective	2	1.62	1.08 to 2.43	0.0		
Surgical approach	Open / laparoscopic mixed	4	1.77	1.26 to 2.49	0.0	0.021	0.106
	Laparoscopic predominant	2	0.84	0.49 to 1.45	0.0		
	Open predominant	1	2.88	1.67 to 4.96	—		
	Open	2	1.50	0.77 to 2.91	53.8		
	Robotic	1	2.72	1.25 to 5.92	—		
Definition detail	Strength + performance + mass	2	1.94	0.32 to 11.95	55.8	0.797	1.000
	Strength only (mass and quality matched out)	1	0.49	0.04 to 5.57	—		
	Strength + mass + muscle-specific strength	1	1.68	1.11 to 2.53	—		
	Mass only	6	1.71	1.12 to 2.62	62.8		

CI, confidence interval; k, number of contributing studies; OR, odds ratio. No interaction test survived Holm correction. The nominally significant surgical-approach comparison is discussed in the text and should not be read as a finding.

Two qualifications must be attached to this result. The first is statistical: the function-inclusive subgroup contains only four cohorts and its own confidence interval crosses unity, so this is an absence of evidence for superiority rather than positive evidence of equivalence. The second is that an equal pooled odds ratio is not the same as equal clinical usefulness. The function-inclusive subgroup was markedly more consistent across cohorts, with an I^2 of 21.6% against 62.8% for the imaging-only subgroup. That is the pattern one would expect if a functional criterion measures something more stable across populations than a computed-tomography threshold whose meaning shifts with the reference chart from which it was derived.

The width of that interval deserves to be stated plainly rather than left to inference. With four cohorts in the function-inclusive arm, the data remain compatible with a function-inclusive odds ratio anywhere between roughly half and roughly twice the imaging-only estimate. A difference of the magnitude that would change practice, were one to exist, could not have been detected by this comparison. No retrospective power calculation is offered, because such calculations conditioned on an observed result are uninformative; the confidence interval itself carries the necessary information and is reported alongside every subgroup estimate in Table 3.

3.6. Within-cohort comparisons of competing definitions

Between-study comparisons confound the definition with everything else that differs between cohorts, including case mix, surgical practice and the era of treatment. Three included cohorts, together with one further cohort that could not enter the pool because it graded morbidity at a lower threshold, applied two or more definitions to the same patients. There the picture was considerably more informative than the pooled subgroup contrast, and it pointed consistently in one direction: muscle mass alone was the weakest of the available criteria, and adding a second dimension improved discrimination. These comparisons are set out in Table 4.

Table 4. Within-cohort comparisons in which two or more sarcopenia definitions were applied to the same patients, so that between-cohort confounding cannot explain the difference between them.

Cohort and source	n	Definitions applied to the same patients	Result
Wu 2025, First Affiliated Hospital of Wenzhou Medical University[27]	1,654	GLIS criterion 1 (low muscle-specific strength alone); criterion 2 (low muscle strength plus low muscle mass); criterion 3 (composite)	Criterion 1 was not associated with severe complications (crude OR 1.075, 0.673 to 1.718). Criterion 2 was (OR 2.050, 1.296 to 3.243). Criterion 3 gave OR 1.676 (1.109 to 2.533).
Ishikawa Prefectural Central Hospital[26,30]	840 and 114	CT skeletal muscle index at six published cut-offs; intramuscular adipose tissue content; low handgrip strength after matching on muscle mass and quality	Low skeletal muscle index was null for severe complications at every cut-off tested (Zhuang OR 0.868, 0.497 to 1.520; Prado 0.804; Martin 0.785; Sakurai 0.92; Iritani 1.33). Poor muscle quality in the same patients was an independent risk factor (adjusted OR 2.260, 1.220 to 4.190). After matching on muscle mass and quality, low grip strength remained associated with infectious complications (17.5% vs 1.8%, p = 0.008), although grade IIIa or above events were too few to separate the groups.
Guo 2025, Fourth Hospital of Hebei Medical University[31]	381	Low skeletal muscle index alone; concurrent low skeletal muscle index and myosteatosis	Low skeletal muscle index alone gave OR 2.72 for severe complications; the concurrent phenotype gave OR 6.67 (2.22 to 12.68).
Sato T, et al. Gastric Cancer. 2016;19(3):1008-15. doi: 10.1007/s10120-015-0554-4	293	Low handgrip strength; low lean body mass	Low handgrip strength predicted morbidity whereas low lean body mass did not. This cohort graded morbidity at Clavien-Dindo grade II or above and so could not enter the primary pool; it is reported narratively.

CT, computed tomography; GLIS, Global Leadership Initiative in Sarcopenia; OR, odds ratio. The fourth cohort is cited in full within the table because it did not contribute to any pooled estimate.

The most instructive of these is the single cohort of 1,654 patients in which three operationalisations of one consensus framework were applied to the same people. Requiring muscle-specific strength alone produced no association with severe complications, requiring low muscle strength together with low muscle mass produced a clear one, and the composite criterion fell between the two[27]. Within one cohort, with case mix, surgeons and era held constant, what the definition requires changes the answer. Data from Matsui and colleagues make the complementary point from the other direction: in 840 patients, low skeletal muscle index was null for severe complications at every one of six published cut-offs, while poor muscle quality assessed as intramuscular adipose tissue content in the same patients was an independent risk factor[30].

A second observation from that same cohort deserves separate emphasis, because it demonstrates the threshold problem directly rather than by inference. In one cohort of 840 patients, six published skeletal muscle index cut-off sets were applied to the same people. All six produced statistically null associations with severe complications, but their point estimates ranged from 0.785 to 1.33[30]. The patients, the surgeons, the era and the outcome definition were identical across those six analyses; only the reference chart differed. Whatever a published threshold represents, it is not a fixed property of the patient.

3.7. Structure of the published evidence

An audit of the overlap structure explains why the primary comparison cannot currently be powered, and why this is a ceiling imposed by the literature rather than a failure of searching. Twelve published reports of preoperative sarcopenia and gastrectomy outcomes were identified from a single department, ten of them using function-inclusive definitions and nested within one of two accrual windows. Two of those twelve, both appearing in the same journal and describing an identical accrual window, could not be distinguished from one another as separate cohorts and are presented as a single entry, so eleven rows appear in Table 5. Under the clustering rule only two of these could enter the primary pool. Every other function-inclusive gastrectomy cohort in the world literature that reports the outcome at grade IIIa or above either sits behind a paywall or grades morbidity at a lower threshold. The audit is presented in Table 5.

3.8. Secondary subgroups and multiplicity

Neither geographical region nor study design modified the association. Asian cohorts gave 1.87 (1.15 to 3.03) and Western cohorts 1.46 (1.00 to 2.12), with an interaction p of 0.429; retrospective cohorts gave 1.77 (1.18 to 2.66) and prospective cohorts 1.62 (1.08 to 2.43), with an interaction p of 0.750. Surgical approach reached nominal significance with a raw interaction p of 0.021, but this should not be reported as a finding. Three of its five levels contain only one or two cohorts, and the laparoscopic-predominant level consists solely of the two reports of Matsui and colleagues, which are also the two null ones. After Holm correction for the five interaction tests performed, no subgroup analysis in this review was significant, the surgical-approach comparison reaching a corrected p of 0.106. The same conclusion followed under the Benjamini-Hochberg procedure. Every subgroup estimate, with its raw and corrected interaction p value, is detailed in Table 3.

Table 5. Audit of serially published reports from a single department, showing why only two could enter the primary pool. Reports that contribute nothing to any pooled estimate are cited in full within the table.

Report	n	Accrual window	Definition family	Disposition in this review
Zhuang CL, et al. Medicine (Baltimore). 2016;95(13):e3164[28]	937	Dec 2008 - Apr 2013	Imaging-mass only	**Included** — primary pool, Wenzhou-early cluster
Lou N, et al. Eur J Surg Oncol. 2017;43(1):188-195. doi: 10.1016/j.ejso.2016.09.006	206	2014 - 2016	Function-inclusive	Excluded — nested within a retained cluster
Zhuang CL, et al. Clin Nutr. 2020;39(7):2301-2310. doi: 10.1016/j.clnu.2019.10.024	883	Aug 2014 - Feb 2018	Function-inclusive	Excluded — nested within a retained cluster
Chen XY, et al. Eur J Surg Oncol. 2019;45(6):1092-1098. doi: 10.1016/j.ejso.2018.09.030	313	2014 - 2017	Function-inclusive	Excluded — nested within a retained cluster
Huang DD, et al. J Cancer. 2020;11(19):5852-5860. doi: 10.7150/jca.49815	880	From Aug 2014	Function-inclusive	Excluded — nested within a retained cluster
Huang DD, et al. Nutrition. 2021;86:111156[33]	419	Aug 2014 - Jun 2019	Function-inclusive	Excluded — nested within a retained cluster
Zou HB, et al. Eur J Surg Oncol. 2021;47(8):1976-1984[34]	135	Dec 2015 - Jun 2017	Function-inclusive	Excluded — nested within a retained cluster
Zhang FM, et al. Front Nutr. 2021;8:709211 (normal body-mass-index subset)[35]	267	Jul 2014 - Dec 2016	Function-inclusive	Excluded — nested within a retained cluster
Huang DD, et al. Eur J Clin Nutr. 2022;76(9):1323-1331[36]	1,359	Aug 2014 - Jun 2019	Function-inclusive	Excluded — nested within a retained cluster
Wang SL, et al. Eur J Surg Oncol. 2025;51(9):110229[37]	1,404	2013 - 2019	Function-inclusive	Excluded — cluster alternate, retained for sensitivity analysis
Wu GF, et al. BMC Cancer. 2025;25(1):679[27]	1,654	Jul 2014 - Dec 2021	Function-inclusive	**Included** — primary pool, Wenzhou-late cluster

Two of these entries, both in Clinical Nutrition with an identical accrual window of August 2014 to February 2018, could not be distinguished as separate cohorts and are presented here as one report. The accrual windows are those recorded during data extraction.

3.9. Sensitivity analyses

The pooled estimate was robust. Leave-one-out removal gave estimates between 1.55 and 1.89, and every one of the ten remained statistically significant, so no single cohort drives the result. Collapsing the Matsui cluster by removing the propensity-matched subsample, so that the institutional rule is applied without the exposure-construct exception described in Section 2.4, gave 1.74 (1.26 to 2.41), which is indistinguishable from the primary estimate. Swapping the alternate report into the Guo cluster gave 1.72 (1.23 to 2.42). Restricting the pool to the five institutionally independent cohorts gave 1.67 (1.07 to 2.61), and restricting it to the eight cohorts with directly reconstructable counts gave 1.96 (1.35 to 2.85). Adding the two pre-specified sarcopenic-obesity cohorts gave 1.85 (1.37 to 2.51). Re-fitting the count-based studies with an exact within-study binomial likelihood changed essentially nothing, giving 1.94 (1.39 to 2.69) against 1.96 (1.35 to 2.85) for the inverse-variance model, with Peto at 1.94 and Mantel-Haenszel at 1.91, so the result does not depend on the normal approximation to the log odds. The complete set of sensitivity analyses is detailed in Table 6.

Table 6. Sensitivity analyses. Every leave-one-out estimate remained statistically significant, and the full range across the ten omissions was 1.55 to 1.89.

Sensitivity analysis	k	Pooled OR	95% CI	p	I^2 (%)
Primary analysis	10	1.71	1.24 to 2.35	0.0012	48.0
Leave-one-out, minimum (omitting Zhuang 2016)	9	1.55	1.13 to 2.14	0.0066	37.2
Leave-one-out, maximum (omitting Matsui 2021)	9	1.89	1.41 to 2.54	< 0.001	29.2
Collapse the Matsui cluster (institutional rule applied without the exposure-construct exception)	9	1.74	1.26 to 2.41	< 0.001	51.1
Cluster swap: Ding 2024 replaces Guo 2025[38]	10	1.72	1.23 to 2.42	0.0015	50.3
Institutionally independent cohorts only	5	1.67	1.07 to 2.61	0.025	28.9
Studies with a reconstructable two-by-two table only	8	1.96	1.35 to 2.85	< 0.001	36.3
Adding the pre-specified sarcopenic-obesity cohorts[39,40]	12	1.85	1.37 to 2.51	< 0.001	46.7
Hartung-Knapp-Sidik-Jonkman adjustment	10	1.71	1.17 to 2.48	0.010	48.0
Restricted maximum likelihood τ^2 estimator	10	1.71	1.23 to 2.36	0.0013	48.0
Binomial-normal generalised linear mixed model (count studies)	8	1.94	1.39 to 2.69	< 0.001	25.1
Peto fixed effect (count studies)	8	1.94	1.48 to 2.53	< 0.001	—
Mantel-Haenszel fixed effect (count studies)	8	1.91	1.46 to 2.50	< 0.001	—
Pool of study-level adjusted estimates	3	1.96	1.27 to 3.02	0.0024	9.5
Pool of adjusted estimates with Hartung-Knapp	3	1.96	0.76 to 5.05	0.093	9.5
Within the function-inclusive subgroup, omitting Wu 2025 (the fragility quoted in Section 3.9)	3	1.42	0.32 to 6.24	0.644	46.6
Within the imaging-mass-only subgroup, omitting Zhuang 2016 (the fragility quoted in Section 3.9)	5	1.50	0.99 to 2.28	0.056	51.3

CI, confidence interval; k, number of contributing studies; OR, odds ratio. The last two rows show the within-subgroup fragility discussed in the text. The two rows labelled as pooling adjusted estimates pool each study's own multivariable estimate rather than the crude estimate; with only three contributing studies this analysis is exploratory and the Hartung-Knapp interval is the honest one to read.

Within the function-inclusive subgroup the estimate was fragile, as the small number of cohorts implies: omitting the largest collapsed it to 1.42 (0.32 to 6.24). Within the imaging-only subgroup, omitting the largest left 1.50 (0.99 to 2.28), which loses significance at the conventional threshold. Both fragilities are recorded in the last two rows of Table 6 rather than buried, because they bear directly on how much weight the primary comparison can carry.

3.10. Study-level adjusted estimates

Only three of the ten cohorts published a multivariable estimate for the primary endpoint, and the three adjusted for materially different variable sets, ranging from a fully enumerated model including stage and comorbidity to one whose adjustment set was not stated. They are therefore presented individually in Table 7 rather than being relied upon as a pooled quantity. All three point in the same direction, with adjusted odds ratios of 4.76, 1.63 and 2.56, each with a confidence interval excluding unity. Their exploratory pooled value, 1.96 (1.27 to 3.02) with an I^2 of 9.5%, is slightly larger than the crude pooled estimate, which is the opposite of what confounding by age, stage and comorbidity would be expected to produce; under the Hartung-Knapp adjustment appropriate to three studies the interval widens to 0.76 to 5.05 and no longer excludes unity. The narrative conclusion that the association is not wholly attributable to measured confounding therefore rests on the consistent direction of three individually adjusted estimates, not on their pooled value.

Table 7. Study-level multivariable estimates for major postoperative complications, presented individually. Only the first three could contribute to a pooled adjusted estimate, and the reason each of the remainder could not is given.

Study	Definition family	Adjusted OR (95% CI)	Variables adjusted for
Fukuda 2016[24]	Function-inclusive	4.76 (1.03 to 24.3)	Multivariable logistic model; the adjustment set was not fully enumerated in the source
Wu 2025[27]	Function-inclusive	1.63 (1.05 to 2.53)	Age, sex, body mass index, American Society of Anesthesiologists grade, TNM stage
Guo 2025[31]	Imaging-mass only	2.56 (1.08 to 6.09)	Age, stage, smoking, alcohol, Charlson comorbidity index, total adipose tissue
Zhuang 2016[28]	Imaging-mass only	3.01 (not reported)	Could not contribute: no confidence interval published (p < 0.001)
Erkul 2022[25]	Function-inclusive	2.73 (1.02 to 7.52)	Could not contribute: the adjusted estimate is for any-grade complications
Sierzega 2019[29]	Imaging-mass only	1.94 (1.08 to 3.48)	Could not contribute: the adjusted estimate is for overall morbidity
Matsui 2021[30]	Imaging-mass only	Not applicable	Could not contribute: low muscle index was modelled univariately only
Matsui 2022, Back 2025, Akmercan 2025[18,26,32]	Both families	Not applicable	Could not contribute: no multivariable model was fitted for this endpoint

CI, confidence interval; OR, odds ratio. The three contributing estimates differ substantially in what they adjust for, so a pooled quantity derived from them does not correspond to a single causal contrast; they are therefore presented individually and discussed narratively, with the pooled value reported in Table 6 retained only as an exploratory summary.

3.11. Small-study effects

With exactly ten contributing cohorts the pre-specified threshold for these tests was met. There was no evidence of funnel asymmetry: the Egger intercept was 0.0054 with a standard error of 1.1211 (p = 0.996), and the Begg-Mazumdar rank correlation gave p = 0.788. Trim-and-fill imputed no missing studies and left the pooled estimate unchanged at 1.71 (1.24 to 2.35). The two most extreme small cohorts point in opposite directions, one strongly positive and one strongly protective, which is why the funnel remains symmetric despite the modest number of studies. These tests are underpowered at ten studies and the correct reading is that no asymmetry was detected, not that publication bias is absent. The contour-enhanced funnel and its trim-and-fill counterpart are detailed in Figure 4.

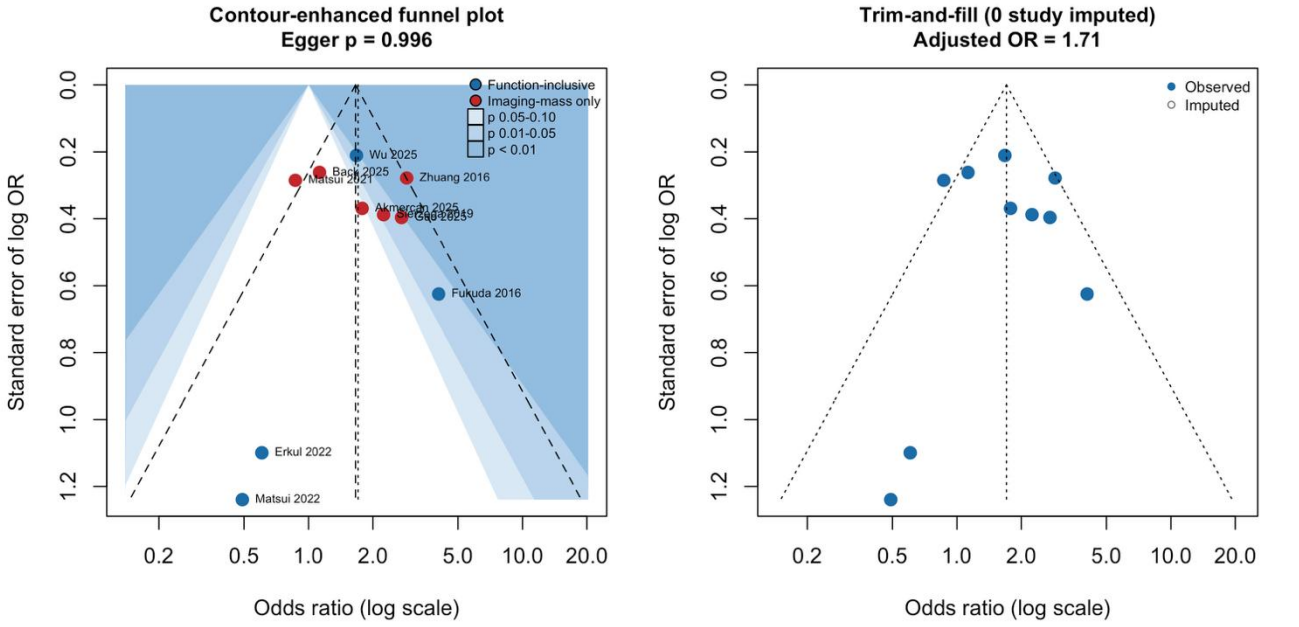

Figure 4. Contour-enhanced funnel plot with study labels and definition-family colouring (left) and the corresponding trim-and-fill funnel (right). No studies were imputed.

3.12. Triangulation against previously published syntheses

Two independent syntheses provide an external check, and they point in different directions on the definition question. A synthesis of nine computed-tomography-mass-only studies comprising 1,614 patients reported an odds ratio of 1.73 (1.14 to 2.63) with an I^2 of 49% for severe complications[20]. That is almost exactly the imaging-only arm of the present review, 1.71 (1.12 to

2.62), obtained from a largely non-overlapping set of studies; the mass-only estimate in this field is therefore well established and externally reproducible. A synthesis organised around consensus-defined sarcopenia across elective abdominal surgery reported, in its gastric subgroup, an odds ratio of 2.66 (1.77 to 4.01) with an I^2 of 0% from six studies[21]. That is substantially higher than the function-inclusive arm of the present review and does not overlap the mass-only estimate. The three estimates are set side by side in Figure 5.

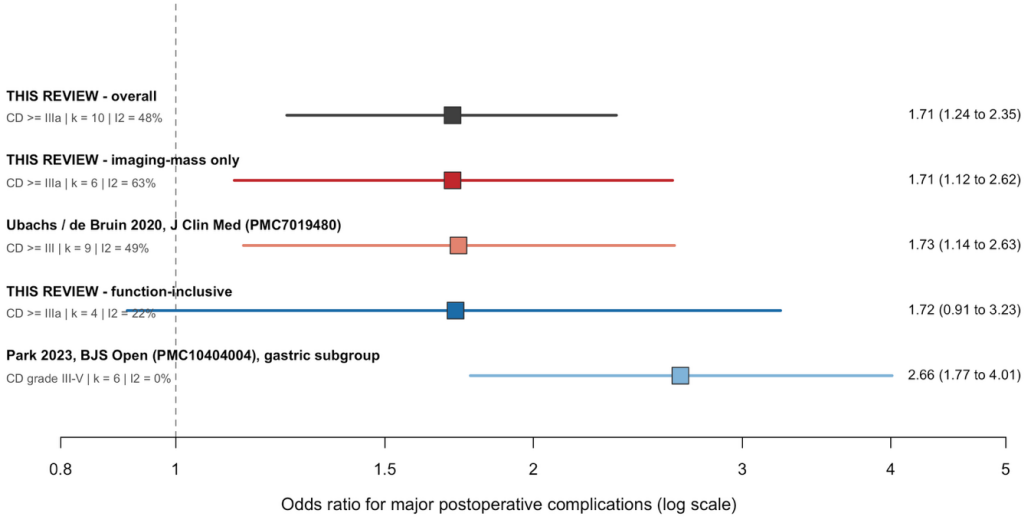

Figure 5. Triangulation of the pooled estimates from this review against two previously published syntheses. The imaging-mass-only estimates agree closely across independent study sets; the external function-inclusive estimate is higher but rests largely on serial reports from a single department. The severity thresholds are not identical across the three sources: this review used Clavien-Dindo grade IIIa or above, the mass-only comparator grade III or above, and the function-inclusive comparator grades III to V, so the comparison is approximate.

The discrepancy has a structural explanation. Four of the six gastric studies in that synthesis are serial reports from the same department, with overlapping accrual windows, and are treated in it as independent cohorts. Its pooled function-inclusive odds ratio of 2.66 is therefore closer to a single-cohort estimate than to a synthesis of six independent cohorts, and its I^2 of 0% reflects that shared provenance rather than genuine consistency across populations.

3.13. Pattern of complications

No secondary outcome met the pre-specified minimum of three cohorts reporting it by sarcopenia status at a common definition, so none was pooled and no summary estimate for any individual complication is offered. The pattern that the individual cohorts describe can nonetheless be stated narratively, and it is clinically relevant because different complications imply different preoperative responses. The available reports point predominantly towards infectious and pulmonary events rather than towards technical failure of the anastomosis. In the Japanese cohort in which grip strength was compared against muscle mass after matching, the association that survived matching was with infectious complications specifically, at 17.5% against 1.8%[26]. Outside gastrectomy, preoperative sarcopenia has been most consistently linked to postoperative pneumonia, which occurred in 31.7% of sarcopenic against 13.3% of non-sarcopenic patients after oesophagectomy[9], and a respiratory formulation of the construct stratified complication risk after lung resection[10]. Prehabilitation trials in this population have likewise reported their clearest effects on pulmonary events[41]. Anastomotic leakage, by contrast, was not consistently associated with sarcopenia in the cohorts examined here, and an earlier synthesis in this field also failed to find such an association. This pattern is a hypothesis rather than a finding: the data are too sparse to characterise it formally, and it is offered only to indicate where a preoperative response might most usefully be directed.

3.14. Certainty of the evidence

Certainty was rated very low for both the overall association and the comparison between definition families, and the domain-by-domain reasoning is detailed in Table 8. The evidence base started at low certainty because it is entirely observational, and was downgraded one further level for imprecision on the basis of a prediction interval that crosses unity. It was not downgraded for inconsistency, indirectness, risk of bias or publication bias, and was not upgraded for magnitude of effect.

Table 8. Certainty of the evidence, assessed with the GRADE approach adapted for prognostic-factor reviews.

GRADE domain	Judgement	Reason
Starting certainty	Low	The whole body of evidence consists of observational cohort studies.
Risk of bias	Not downgraded further	Eight of ten cohorts were rated as raising some concerns and two as high risk; omitting both high-risk cohorts did not change the estimate.
Inconsistency	Not downgraded	I^2 was 48.0%, below the pre-specified 75% threshold, and the subgroup confidence intervals overlapped.
Indirectness	Not downgraded	Population, exposure and outcome were all directly relevant to the question posed.
Imprecision	Downgraded one level	The 95% prediction interval, 0.73 to 3.99, crosses unity, and the Hartung-Knapp interval narrows to 1.17 to 2.48.
Publication bias	Not downgraded	No funnel asymmetry was detected and trim-and-fill imputed no missing studies.
Large effect	Not upgraded	A pooled odds ratio of 1.71 is a modest association.
Overall association	**Very low**	—
Difference between definition families	**Very low**	The null interaction rests on only four function-inclusive cohorts and cannot support a claim of equivalence.

4. Discussion

4.1. Principal findings

Across ten cohorts and 4,883 patients, preoperative sarcopenia was associated with roughly a seventy per cent increase in the odds of a major postoperative complication after gastrectomy for gastric cancer. The association survived every sensitivity analysis attempted: leave-one-out removal, the choice of report within an overlapping cluster, restriction to institutionally independent cohorts, the estimator of between-study variance, and replacement of the normal approximation to the log odds with an exact binomial likelihood. Pooling each study's own multivariable estimate gave a slightly larger association than the crude pool, which is the opposite of what confounding by age, stage and comorbidity would produce and therefore argues that the association is not merely a reflection of those variables.

The question the title asks received a negative answer. A sarcopenia definition that incorporates muscle strength or physical performance did not produce a larger pooled association than one based on imaging muscle mass alone; the two estimates differed in the second decimal place and the interaction test was effectively null. Meta-regression on definition family, and on a graded count of the constructs each definition required, explained none of the between-study variance. Taken at face value, this licenses the pragmatic behaviour the field has already adopted, which is to read muscle area from a staging computed tomogram that has already been acquired.

That reading would nonetheless be too quick, for three reasons visible in these data. The first is statistical power: four cohorts cannot establish equivalence, and the function-inclusive confidence interval, 0.91 to 3.23, is compatible with associations ranging from mildly protective to strongly harmful. The second is consistency: although the pooled point estimates coincided, the function-inclusive estimate was far more homogeneous across cohorts, with an I^2 of 21.6% against 62.8%. A criterion anchored to a measured functional capacity plausibly travels between populations better than a threshold inherited from a reference chart that may not describe the patients to whom it is applied. The third, and the most persuasive, is the within-cohort evidence.

4.2. What the within-cohort comparisons add

Where the same patients were classified under competing definitions, the definitions did not perform equally. In a single cohort of 1,654 patients, one operationalisation of a consensus framework produced no association with severe complications, another produced a clear one, and a third fell between them[27]. In the 840-patient cohort of Matsui and colleagues, low skeletal muscle index was null at all six published thresholds tested, while muscle quality measured in the same patients was an independent predictor[30]. In a third cohort, the combination of low muscle mass with myosteatosis carried more than twice the odds of low muscle mass alone[31]. In a fourth, which could not enter the pool because it graded morbidity at a lower threshold, low handgrip strength predicted morbidity while low lean body mass did not.

These comparisons are internally controlled, and they converge on a claim that the pooled subgroup analysis is not powerful enough to detect: muscle mass alone is the weakest single criterion, and what improves prediction within a cohort is the addition of a second dimension, whether that dimension is strength or muscle quality, rather than the particular consensus label attached to the resulting definition. This reframes the question. The useful distinction may not be between named consensus definitions and imaging surrogates, but between one-dimensional and multi-dimensional characterisations of muscle. Evidence from adjacent fields is consistent with this. Low skeletal muscle radiodensity, not low muscle area, was the best predictor of major complications after gastrointestinal cancer surgery in one prospective cohort[42]; myosteatosis and intermuscular adiposity were independent prognostic factors in a two-centre development and validation study in colorectal cancer[43]; muscle attenuation rather than skeletal muscle index was the independent prognostic factor in overweight patients with gastric cancer[44]; and in oesophageal adenocarcinoma it was myosteatosis rather than sarcopenia that was associated with postoperative complications and survival[45].

Where two lines of evidence within one review disagree, the reader is entitled to know which the authors believe. We give precedence to the within-cohort comparisons, and we do so on grounds of study design rather than of preference. The pooled subgroup contrast compares definitions across cohorts that differ in country, era, case mix, surgical approach and threshold set, with four studies in one arm and an interval spanning a factor of three and a half; the within-cohort comparisons hold every one of those variables constant and agree with one another. A between-study interaction test of that width is close to uninformative, whereas four internally controlled comparisons converging on the same ordering are not. The honest summary is therefore that the pooled analysis failed to detect a difference it was not powerful enough to detect, and that the better-designed comparisons available within this evidence base indicate that a one-dimensional assessment of muscle is the weakest option. We have retained the interrogative title because the question it poses is the contribution of the review and because the answer is genuinely two-part; the Abstract states both parts.

The functional dimension carries independent information in other surgical settings too. Handgrip strength predicted 30-day complications after cardiac surgery in patients aged sixty and over, with an area under the receiver operating characteristic curve of 0.765 for grip recovery[46]; low handgrip strength was associated with major complications after hepatectomy for primary liver cancer, at 40.3% against 24.6%[47]; and probable sarcopenia defined by a screening questionnaire plus low grip strength, when overlapping with malnutrition, independently predicted 30-day readmission after major oncological surgery[48]. Within the gastrectomy literature itself, work from the same department that dominates the function-inclusive evidence has repeatedly found that adding strength or gait speed to a mass measurement improves prediction[33,35,49], that different consensus frameworks applied to the same patients give different prognostic performance[50], and that a grip-strength-based operationalisation of a malnutrition framework outperforms a muscle-mass-based one in the same cohort[36].

4.3. Comparison with previously published syntheses

The imaging-only estimate from this review agrees closely with an independently assembled synthesis. A review of nine computed-tomography studies reported 1.73 (1.14 to 2.63) for severe complications, against 1.71 (1.12 to 2.62) here[20]. The two study sets were assembled from different search windows and are not identical, but we did not re-extract the comparator's study list and therefore cannot enumerate how many primary studies are common to both; the agreement should be read as supportive rather than as fully independent replication. The most recent and largest synthesis in the field, restricted by design to the skeletal muscle mass index, likewise reported significantly higher rates of severe complications in sarcopenic patients[19]. The present review therefore does not overturn that literature; it clarifies what that literature is an estimate of, and shows that it is stable.

The one previous synthesis organised around consensus-defined sarcopenia reported a considerably larger gastric-subgroup association, 2.66 (1.77 to 4.01), with an I^2 of 0%[21]. The separation between that estimate and the two mass-only estimates is visible in Figure 5, and the present analysis suggests why it arises. Four of that synthesis's six gastric studies are serial reports from one department with overlapping accrual windows, entered as independent cohorts. A pooled estimate assembled that way approximates a single-cohort estimate with artificially narrow confidence limits, and an I^2 of zero in such a set reflects shared provenance rather than agreement between populations. The audit in Table 5 shows the scale of the problem: at least twelve reports, and by our own search of that department's output considerably more, describe patients drawn from two accrual windows in one hospital. This is not a criticism of those investigators, whose serial reporting of a well-characterised prospective cohort is legitimate and valuable. It is a criticism of

how syntheses in this field have handled that output. Any future meta-analysis of sarcopenia in gastric cancer that does not resolve cohorts by institution and accrual window will overstate both the size of the effect and the consistency of the evidence, and the correction is straightforward to apply.

4.4. Heterogeneity

Heterogeneity was moderate overall, at an I^2 of 48.0%, and it was concentrated in the imaging-only subgroup, at 62.8%, rather than in the function-inclusive subgroup, at 21.6%. None of the pre-specified moderators explained it. Definition family, publication year, sample size and the number of constructs required were all null in meta-regression, with an R^2 of zero in every model that included definition family.

The most plausible residual source is the threshold problem within the imaging-only group. Six cohorts all measured the skeletal muscle index at the third lumbar vertebra, yet applied four different threshold sets derived from different reference populations. A patient is not sarcopenic or not sarcopenic in any absolute sense under such a scheme; the label depends on which chart was consulted. The evidence that this matters is direct and internal to these data: in one cohort, six published cut-offs applied to the same 840 patients all produced null associations but with point estimates ranging from 0.785 to 1.33[30]. Outside surgery, the same phenomenon is well documented, with prevalence in identical populations varying several-fold according to the criterion applied[12-15]. Automated and deep-learning segmentation now makes the measurement itself highly reproducible, with correlations against manual measurement of 0.776 or better for skeletal muscle and intraclass coefficients above 0.98 in validation work[51,52], which means the remaining variability is attributable to the threshold rather than to the measurement. Standardising the threshold, not the software, is where the field's effort now belongs.

We did not test this proposition formally, and we state that rather than leaving it implied. A subgroup analysis or meta-regression on threshold set was not pre-specified and was not performed, because the six imaging-only cohorts distribute across four threshold sets in a pattern of three, one, one and one; a moderator with three singleton levels cannot be estimated with any useful precision, and fitting it would have produced a coefficient that looked like evidence and was not. The explanation offered here therefore remains a hypothesis supported by within-cohort data rather than a tested moderator, and testing it will require either individual patient data or a substantially larger set of independent cohorts.

Two further cautions about the heterogeneity statistics belong here. The confidence interval around the I^2 of 48.0% runs from 0.0% to 74.9%, so these data are also compatible with no between-study heterogeneity at all, and the point estimate should not be treated as a stable description of the evidence. The same applies with greater force to the 95% prediction interval of 0.73 to 3.99. A prediction interval is a function of an estimated between-study variance, and at ten studies with a DerSimonian-Laird estimator that variance is itself poorly determined; such intervals are known to be unreliable at small numbers of studies. We used the interval to downgrade certainty for imprecision because it points in the conservative direction, but it should not be read as a precise statement about a future setting. The corresponding interval under the restricted-maximum-likelihood specification was not computed in the analysis and is therefore not reported here.

4.5. Limitations

First, and most importantly, the primary comparison is underpowered. Only four cohorts applied a function-inclusive definition and reported the outcome at the required severity threshold, and the resulting confidence interval, 0.91 to 3.23, includes both a protective and a strongly harmful association. Nothing in this review should be read as evidence that the two definition families are equivalent; what is reported is that no difference was detected. Within that subgroup the estimate was also fragile, collapsing to 1.42 (0.32 to 6.24) when the largest cohort was removed.

Second, and closely related, nine eligible reports could not be retrieved because the full text sat behind a subscription paywall and the event counts could not be recovered from the abstract. Their identities, definition families and reported directions of effect are set out in Table 9. Four of the nine used function-inclusive definitions, which is the same order of magnitude as the arm they would have joined, so this is not a peripheral limitation but a mechanism that could plausibly have changed the primary comparison. The direction matters and can be stated: of those four, three report a positive association and one is null, so their inclusion would more probably have raised than lowered the function-inclusive pooled estimate. It follows that the null primary comparison reported here may be an artefact of retrieval failure rather than a property of the literature, and it should be read with that possibility explicitly in mind. One of the four is a within-cohort comparison of a grip-strength criterion with a muscle-mass criterion in more than 1,300 patients, for which only the two p values are public, and it is the single most valuable missing item. Corresponding authors were not contacted for the missing counts; doing so is the first step any update of this review should take.

Table 9. The nine eligible reports whose full text could not be obtained, with the direction of effect stated in each abstract. Four used function-inclusive definitions, which is the arm most vulnerable to their absence.

Report not retrieved	Definition family	Endpoint reported	Direction of effect stated in the abstract
Gupta 2026, J Surg Oncol	Function-inclusive (muscle index plus grip strength)	Grade IIIa or above	Strongly positive: combined weakness independently predicted major complications
Kim 2023, Eur Radiol	Function-inclusive (consensus criteria)	Grade III or above	Positive: univariate odds ratio 3.77 (p = 0.036), no confidence interval published
Huang 2022, Eur J Clin Nutr	Function-inclusive (grip-strength and muscle-mass operationalisations)	Severe complications	Mixed and directly relevant: the grip-strength criterion was associated (p = 0.046), the muscle-mass criterion was not (p = 0.270)
Wagh 2024, Cancer Treat Res Commun	Function-inclusive (grip strength plus muscle index)	Grade 3 or above	Null: no significant association (p = 0.857)
Lee 2026, J Gastric Cancer	Imaging-mass only	Grade II or above	Positive for the combined frailty-and-sarcopenia phenotype; no estimate for sarcopenia alone
Tuncer 2023, Acta Chir Belg	Imaging-mass only	Clavien-Dindo grade, threshold not stated	Positive: complication grade higher in sarcopenic patients (p = 0.006)
Sinduja 2022, J Gastrointest Cancer[17]	Imaging-mass only	Clavien-Dindo grade, threshold not stated	Null: no association with any postoperative complication
Shi 2019, Ann Nutr Metab	Imaging-mass only	Clavien-Dindo grade, threshold not stated	Not determinable: complications reported only in aggregate
Alnimri 2021, ANZ J Surg	Imaging-mass only	Major complications, threshold not stated	Not usable: the published percentages are proportions within the complicated patients rather than incidences

Reports are cited in full within the table where they do not appear in the reference list, since none contributed to any pooled estimate. Of the four function-inclusive reports, three point towards a positive association and one is null, so their inclusion would more probably have raised than lowered the function-inclusive pooled estimate.

Third, the evidence base is entirely observational and predominantly retrospective, with confounding rated the weakest ROBINS-E domain and only 30% of cohorts rated at low risk in that domain. Retrospective cohorts assembled from patients with an interpretable preoperative computed tomogram exclude those without one in a manner unlikely to be random. Two cohorts were rated at high risk of bias overall. Although the adjusted-estimate pool argues against confounding as the whole explanation, it rests on only three cohorts and loses significance under the Hartung-Knapp adjustment.

Fourth, screening was carried out by a single reviewer against English-language records in two databases. This departs from the dual-reviewer standard and cannot be remedied retrospectively. It carries more weight in this review than it would in most, for two reasons: the primary comparison rests on four cohorts in one arm, so a single missed study could materially alter it, and the identification of overlapping cohorts in this literature requires judgement about accrual windows and institutional provenance rather than mechanical application of eligibility criteria. This limitation is stated in the Abstract as well as here, so that a reader relying on the Abstract alone is aware of it. Eligible reports published in languages other than English may also have been missed.

4.6. Transportability of published thresholds

The cohorts pooled here were operated on in China, Japan, Turkey, Poland and Finland. None was from South or South-East Asia, and this bears directly on how the review's most practical recommendation should be applied. Skeletal muscle index thresholds are not physiological constants; they are percentile cut-offs derived from a specific reference population, and the four sets in use among the six imaging-only cohorts here originate from different populations. Body composition differs systematically between East Asian, South-East Asian and European populations at the same body mass index, so a threshold derived in one is not guaranteed to identify the same risk stratum in another. The internal evidence that this matters is direct: applying six published cut-off sets to one cohort of 840 patients moved the point estimate between 0.785 and 1.33[30], and outside surgery the prevalence of sarcopenia in a single population has been shown to vary several-fold according to which criterion is applied[12,13].

Two consequences follow for readers practising outside the populations represented here. The first is that a published skeletal muscle index threshold should be adopted with the reference population named, and its performance should be checked locally before it is used to make individual decisions; a locally derived percentile, even a crude one, may classify patients more sensibly than an imported absolute value. The second is that this is a specific practical advantage of the functional dimension. Grip strength is measured directly in the patient in front of the clinician, and although the cut-off used to call it low is also population-derived, the measurement itself does not depend on an external reference chart in the way that a muscle-area threshold does. Where the transportability of imaging thresholds is genuinely uncertain, that is an argument for adding the functional measurement rather than for abandoning assessment.

4.7. Clinical implications

For the surgeon planning an operation today, the practical message is that a preoperative assessment of muscle carries real prognostic information and that the cheapest available version of it is not obviously inferior. Reading the skeletal muscle index from a staging computed tomogram that has already been acquired costs nothing additional and identifies a group with substantially raised odds of a major complication. That is worth doing, and the growing availability of validated automated segmentation makes it feasible at scale[51,52].

The within-cohort evidence nonetheless suggests that a second dimension should be added where it is practicable, and the cheapest such dimension is a hand dynamometer in the pre-admission clinic. It requires no additional imaging, no radiation and no radiologist. It is not, however, free of requirements: it needs a calibrated instrument, a trained assessor, a standardised protocol specifying posture, elbow position and the number of attempts, and a patient able to comply. Grip strength is affected by hand dominance, by arthritis and by pain, and a value obtained without attention to these will not carry the information reported here. Where the staging scan is already being read, muscle radiodensity can be extracted at the same time at no additional cost, and several lines of evidence suggest it may carry more information than muscle area[42-44].

The clinical value of identifying these patients depends on there being something to offer them, and here the evidence has moved, though less decisively than the individual trial results suggest. A multicentre randomised trial of supervised multimodal prehabilitation in older patients with frailty and gastric cancer reduced complications from 28.7% to 17.2%[53]. A randomised trial of exercise and nutrition prehabilitation before gastrectomy attenuated the loss of lean body mass, with a mean difference of 0.4 kg[54]. Supervised home-based prehabilitation during neoadjuvant chemotherapy reduced postoperative pneumonia with a number needed to treat of nine[41], and a three-week programme in frail older patients found no difference in its primary composite complication index but a lower incidence of severe complications, at 11.1% against 25.9%[55]. These trials differ substantially in population, in the content, duration and supervision of the intervention, and in their primary endpoints, and at least one reported a null primary result alongside a positive secondary one. Critically, none of them selected or randomised patients on the basis of sarcopenia status. They therefore establish that prehabilitation benefits frail or unselected patients undergoing gastrectomy; they do not establish that reversing sarcopenia specifically improves outcome. Sarcopenia is best described on this evidence as a marker for which a plausible but not yet directly demonstrated intervention exists, and the case for measuring it rests on that more modest claim.

Drawing these strands together, Table 10 sets out a preoperative pathway with each step graded against the evidence assembled here rather than asserted. It is offered as a structure for local protocol development and not as a validated instrument: no step in it has been prospectively tested as a package, and the strength attaching to steps three and four is derived from within-cohort comparisons rather than from the pooled analysis.

Two further considerations follow. The apparent attenuation of the association in laparoscopic-predominant cohorts did not survive correction for multiplicity and rests on two reports from one institution, so it cannot be treated as a finding; it is nonetheless a hypothesis worth testing prospectively, and it is consistent with a report that laparoscopic radical gastrectomy was a protective factor for clinical outcomes specifically in patients with sarcopenia[56]. And sarcopenia does not end at discharge: a substantial proportion of patients who are not sarcopenic before operation become so within a year, with markedly worse survival when they do[57], and perioperative body-composition change independently predicts readmission[58]. Preoperative measurement is best understood as the first point in a trajectory rather than as a single gate.

4.8. Implications for research

Three priorities follow directly from these data. The field needs prospective gastrectomy cohorts that measure muscle mass and muscle function in the same patients and report the outcome at a harmonised severity threshold, because that is the only design that can settle the definition question; the within-cohort comparisons summarised in Table 4 show how much more informative such a design is than another between-study synthesis would be. Second, the field needs agreement on threshold reporting, since the measurement is now reproducible and the threshold is not. We are not in a position to nominate a single set as correct, because no included cohort compared threshold sets against a clinical outcome in a way that would identify a best performer, and asserting one on the strength of this review would be exactly the kind of unsupported recommendation we have criticised elsewhere. What we do recommend is a reporting convention that costs nothing: that

every primary study name the threshold set it applies and its source population, publish the two-by-two table so that others can re-analyse under a different threshold, and where feasible report the association under at least two published sets, as one included cohort did across six. Third, future syntheses should resolve serially reported cohorts by institution and accrual window before pooling, since failing to do so has already produced at least one published estimate that is substantially larger and apparently more consistent than the underlying evidence supports.

Table 10. A proposed preoperative pathway, with the strength of each step graded against the evidence assembled in this review rather than asserted. No step in this pathway has been prospectively validated as a package.

Step	Action	Who	Strength of recommendation on the present evidence
1. Identify	Assess every patient scheduled for gastrectomy with curative intent. Do not restrict assessment to older or visibly frail patients: in the pooled cohorts, prevalence exceeded 40% and included patients of normal body mass index.	Surgical team at the point of listing	Reasonable; supported by a consistent association across ten cohorts
2. Measure muscle mass	Read the skeletal muscle index at the third lumbar vertebra from the staging computed tomogram that has already been acquired. Record the threshold set used, by name, in the operation note or database.	Radiology or the surgical team, using validated automated segmentation where available	Reasonable; no additional cost, and the pooled association is externally reproducible
3. Add a second dimension	Measure handgrip strength at the pre-admission clinic with a calibrated dynamometer and a standardised protocol, or extract muscle radiodensity from the same computed tomogram.	Pre-admission clinic nurse, or radiology	Weaker but supported by internally controlled within-cohort comparisons rather than by the pooled analysis
4. Interpret	Treat low muscle mass together with either low grip strength or low muscle radiodensity as the higher-risk phenotype. Do not treat a single normal muscle index as reassurance, since it was the weakest single criterion in every within-cohort comparison examined.	Surgical team	Weaker; based on four within-cohort comparisons, none of which was powered for the primary endpoint
5. Act	Refer to a multimodal prehabilitation pathway combining supervised exercise and nutritional support where one exists, and ensure the anaesthetic and critical-care plan reflects the raised risk.	Multidisciplinary team	Indirect: prehabilitation trials improved outcomes in frail or unselected patients, not in patients selected by sarcopenia status
6. Re-measure	Repeat the assessment during follow-up. A substantial minority of patients who are not sarcopenic before operation become so within the first postoperative year, with worse survival when they do.	Follow-up clinic	Reasonable for prognosis; no evidence yet that intervening on it changes outcome

4.9. Relevance to practice in the region

Gastric cancer is a disease of major importance across Asia, and the constraints under which it is treated vary widely between and within countries in the region. The pragmatic conclusion of this review is well matched to those constraints: the most informative single measurement identified here is one that can be extracted from a staging computed tomogram that has already been acquired and paid for, requiring no additional test, no additional visit and no additional radiation exposure. Where automated segmentation is available it makes this feasible at scale[51,52], and where it is not, a manual measurement at a single vertebral level is within the competence of any radiology department. The second dimension we recommend adding, handgrip strength, is among the least expensive measurements in clinical medicine. Set against the cost of managing a Clavien-Dindo grade IIIa complication and the demonstrated consequences of such a complication for long-term survival, the economic case for routine preoperative assessment in a resource-constrained setting is at least as strong as in a well-resourced one. The principal caveat is the one set out in Section 4.6: the thresholds available in the published literature were derived elsewhere, and their local performance should be examined before they are relied upon for individual decisions.

5. Conclusion

Preoperative sarcopenia was associated with major postoperative complications after gastrectomy for gastric cancer, with a pooled odds ratio of 1.71 (95% confidence interval 1.24 to 2.35) across ten cohorts and 4,883 patients. The association survived every sensitivity analysis performed and was not explained by measured confounding. Certainty was nonetheless rated very low, because the evidence base is entirely observational and the prediction interval crosses unity.

A definition incorporating muscle function did not yield a larger pooled association than one based on imaging-derived muscle mass alone, at 1.72 (0.91 to 3.23) against 1.71 (1.12 to 2.62) with a test for subgroup difference of p = 0.997. That comparison rests on four cohorts, four further function-inclusive reports could not be retrieved and three of those four report a positive association, and screening was performed by a single reviewer; it must therefore be read as no difference detected rather than as equivalence. Where competing definitions were applied to the same patients, so that between-cohort confounding could not operate, the ordering was consistent: a one-dimensional assessment of muscle was the weakest criterion, and adding a second dimension, whether muscle strength or muscle quality, improved discrimination. We give precedence to those internally controlled comparisons over the underpowered between-study contrast, and the practical recommendation follows from them.

Reading the skeletal muscle index from a staging computed tomogram that has already been acquired is a defensible and essentially costless way to identify patients at raised risk of major morbidity, provided the threshold set is named and its local performance examined. Where a second dimension can be added cheaply, by hand dynamometry in the pre-admission clinic or by extracting muscle radiodensity from the same scan, the evidence assembled here indicates that it should be. Separately, the function-inclusive literature in this field is far less independent than it appears, being dominated by serially published reports from a single department; syntheses that treat those reports as independent overstate both the magnitude and the consistency of the association. Future reviews should resolve cohorts by institution and accrual window before pooling, and future primary studies should measure mass and function in the same patients, name their threshold set, publish their two-by-two tables, and report morbidity at a harmonised Clavien-Dindo threshold.

Ethical approval statement

This study was a systematic review and meta-analysis of previously published, de-identified aggregate data. No new patient data were collected and no individual patient records were accessed, so institutional ethical approval and informed consent were not required.

Acknowledgments

The authors thank the investigators of the primary cohorts synthesised here, whose published data made this analysis possible.

Funding statement

This study did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Author contribution statement

Conception and design: IPO, GIBB. Literature search and study selection: IPO. Data extraction and risk-of-bias assessment: IPO, NSA. Statistical analysis: IPO, GIBB. Interpretation of results: IPO, GIBB, NSA. Manuscript writing: IPO. Critical revision for important intellectual content: GIBB, NSA. All authors read and approved the final manuscript.

Data availability statement

All data analysed in this review were extracted from the published reports cited in the reference list. The extraction dataset and the analysis outputs are available from the corresponding author on reasonable request.

Generative artificial intelligence disclosure

The authors declare that no generative artificial intelligence tools were used in the design, conduct, analysis, or interpretation of the study or in drafting the scientific content of this manuscript.

6. References

1. Morgan E, Arnold M, Camargo MC, et al. The current and future incidence and mortality of gastric cancer in 185 countries, 2020-40: a population-based modelling study. EClinicalMedicine. 2022; 47: 101404. doi: 10.1016/j.eclinm.2022.101404.
2. Xu Y, Xia C, Wang J, et al. Divergent trends in the burden of esophageal, gastric, and liver cancers in China. J Natl Cancer Cent. 2025; 5(3): 306-312. doi: 10.1016/j.jncc.2025.05.001.
3. Kim JH, Kim WK, Kim J, et al. Spatio-temporal changes in clusters of gastric cancer incidence: the impact of nationwide cancer control programs in South Korea. PLoS One. 2026; 21(6): e0349384. doi: 10.1371/journal.pone.0349384.
4. Son SY, Hur H, Hyung WJ, et al. Laparoscopic vs open distal gastrectomy for locally advanced gastric cancer: 5-year outcomes of the KLASS-02 randomized clinical trial. JAMA Surg. 2022; 157(10): 879-886. doi: 10.1001/jamasurg.2022.2749.
5. Etoh T, Ohyama T, Sakuramoto S, et al. Five-year survival outcomes of laparoscopy-assisted vs open distal gastrectomy for advanced gastric cancer: the JLSSG0901 randomized clinical trial. JAMA Surg. 2023; 158(5): 445-454. doi: 10.1001/jamasurg.2023.0096.
6. Markar SR, Visser MR, van der Veen A, et al. Evolution in laparoscopic gastrectomy from a randomized controlled trial through national clinical practice. Ann Surg. 2024; 279(3): 394-401. doi: 10.1097/SLA.0000000000006162.
7. Obana A, Iwasaki K, Suwa T. Impact of postoperative complications on gastric cancer survival. Surgery. 2025; 178: 108873. doi: 10.1016/j.surg.2024.09.031.
8. Kudou K, Ota M, Ogaki K, et al. Risk factors and prognostic significance of severe postoperative complications after curative resection for remnant gastric cancer: a multicenter retrospective study. J Surg Oncol. 2025; 132(6): 1155-1162. doi: 10.1002/jso.70094.
9. Watanabe Y, Iida M, Nishiyama M, et al. Preoperative sarcopenia predicts complications and non-cancer specific mortality in esophageal cancer surgery. Dis Esophagus. 2025; 38(4): doaf056. doi: 10.1093/dote/doaf056.
10. Sun C, Hirata Y, Kawahara T, et al. Diagnosis of respiratory sarcopenia for stratifying postoperative risk in non-small cell lung cancer. JAMA Surg. 2025; 160(1): 66-73. doi: 10.1001/jamasurg.2024.4800.
11. Dong L, Wen F, Qin LM, et al. Associations between preoperative frailty and major postoperative complications in older surgical patients. J Clin Anesth. 2026; 109: 112099. doi: 10.1016/j.jclinane.2025.112099.
12. Liu X, Hou L, Zhao W, et al. The comparison of sarcopenia diagnostic criteria using AWGS 2019 with the other five criteria in West China. Gerontology. 2021; 67(4): 386-396. doi: 10.1159/000513247.
13. Chiu WC, Kao TW, Peng TC. Prevalence of sarcopenia in Asian older adults: a comparison of nine diagnostic criteria across different regions. Exp Gerontol. 2025; 202: 112721. doi: 10.1016/j.exger.2025.112721.
14. Liu Q, Wang M, Yang Q, et al. Comparison of seven sarcopenia screening tools in older type 2 diabetes patients using four diagnostic criteria. J Nutr Health Aging. 2026; 30(1): 100732. doi: 10.1016/j.jnha.2025.100732.
15. Belfield AE, Wilkinson TJ, Henson J, et al. Sarcopenia prevalence using handgrip strength or chair stand performance in adults living with type 2 diabetes mellitus. Age Ageing. 2024; 53(5): afae090. doi: 10.1093/ageing/afae090.
16. Murnane LC, Forsyth AK, Koukounaras J, et al. Changes in components of sarcopenia diagnostic criteria throughout the surgical treatment of oesophagogastric cancer surgery: a prospective longitudinal study. J Hum Nutr Diet. 2026; 39(1): e70205. doi: 10.1111/jhn.70205.
17. Sinduja R, Anandhi A, Sureshkumar S, et al. Role of sarcopenia in predicting the postoperative morbidity and perioperative mortality in patients undergoing elective surgery for gastric cancer. J Gastrointest Cancer. 2022; 53(4): 939-947. doi: 10.1007/s12029-021-00715-w.
18. Back J, Järvinen T, Sallinen V, et al. Sarcopenia predicts poor long-term survival but not postoperative complications in gastric cancer surgery: an 18-year retrospective cohort study. World J Surg Oncol. 2025; 24(1): 22. doi: 10.1186/s12957-025-04120-6.
19. Liu C, Li Y, Xu Y, et al. The impact of preoperative skeletal muscle mass index-defined sarcopenia on postoperative complications and survival in gastric cancer: an updated meta-analysis. Eur J Surg Oncol. 2025; 51(3): 109569. doi: 10.1016/j.ejso.2024.109569.
20. Borggreve AS, den Boer RB, van Boxel GI, et al. The predictive value of low muscle mass as measured on CT scans for postoperative complications and mortality in gastric cancer patients: a systematic review and meta-analysis. J Clin Med. 2020; 9(1): 199. doi: 10.3390/jcm9010199.
21. Park B, Bhat S, Xia W, et al. Consensus-defined sarcopenia predicts adverse outcomes after elective abdominal surgery: meta-analysis. BJS Open. 2023; 7(4): zrad065. doi: 10.1093/bjsopen/zrad065.
22. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71. doi: 10.1136/bmj.n71.
23. Higgins JPT, Morgan RL, Rooney AA, et al. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environ Int. 2024; 186: 108602. doi: 10.1016/j.envint.2024.108602.
24. Fukuda Y, Yamamoto K, Hirao M, et al. Sarcopenia is associated with severe postoperative complications in elderly gastric cancer patients undergoing gastrectomy. Gastric Cancer. 2016; 19(3): 986-93. doi: 10.1007/s10120-015-0546-4.
25. Erkul O, Cekic AB, Cansu A, et al. Effects of sarcopenia on postoperative outcomes in patients who underwent gastrectomy for gastric cancer. J Surg Res. 2022; 274: 196-206. doi: 10.1016/j.jss.2021.12.051.
26. Matsui R, Inaki N, Tsuji T, et al. Impact of preoperative handgrip strength on postoperative outcome after radical gastrectomy for gastric cancer patients. J Clin Med. 2022; 11(23): 7129. doi: 10.3390/jcm11237129.
27. Wu GF, He CH, Xi WT, et al. Sarcopenia defined by the global leadership initiative in sarcopenia (GLIS) consensus predicts adverse postoperative outcomes in patients undergoing radical gastrectomy for gastric cancer: analysis from a prospective cohort study. BMC Cancer. 2025; 25(1): 679. doi: 10.1186/s12885-025-13967-7.
28. Zhuang CL, Huang DD, Pang WY, et al. Sarcopenia is an independent predictor of severe postoperative complications and long-term survival after radical gastrectomy for gastric cancer: analysis from a large-scale cohort. Medicine (Baltimore). 2016; 95(13): e3164. doi: 10.1097/MD.0000000000003164.
29. Sierzega M, Chrzan R, Wiktorowicz M, et al. Prognostic and predictive implications of sarcopenia in western patients undergoing gastric resections for carcinoma of the stomach. J Surg Oncol. 2019; 120(3): 473-482. doi: 10.1002/jso.25509.
30. Matsui R, Inaki N, Tsuji T. Impact of preoperative muscle quality on postoperative severe complications after radical gastrectomy for gastric cancer patients. Ann Gastroenterol Surg. 2021; 5(4): 510-518. doi: 10.1002/ags3.12452.
31. Guo H, Chen S, Zheng T, et al. Association of low skeletal muscle mass and radiodensity with clinical outcomes in patients undergoing robotic radical gastric cancer surgery: a population-based retrospective cohort study. BMC Cancer. 2025; 25(1): 741. doi: 10.1186/s12885-025-14156-2.
32. Akmercan A, Kıvılcım Uprak T, Buğdaycı O, et al. Computed tomography defined body composition may predict postoperative outcomes and prognosis following gastric cancer surgery. Turk J Surg. 2025; 41(3): 227-234. doi: 10.47717/turkjsurg.2025.6744.
33. Huang DD, Cai HY, Wang WB, et al. Measurement of muscle quantity/quality has additional predictive value for postoperative complications and long-term survival after gastrectomy for gastric cancer in patients with probable sarcopenia as defined by the new EWGSOP2 consensus: analysis from a large-scale prospective study. Nutrition. 2021; 86: 111156. doi: 10.1016/j.nut.2021.111156.
34. Zou HB, Yan XL, Dong WX, et al. Sarcopenia is a predictive factor of poor quality of life and prognosis in patients after radical gastrectomy. Eur J Surg Oncol. 2021; 47(8): 1976-1984. doi: 10.1016/j.ejso.2021.03.004.
35. Zhang FM, Zhang XZ, Shi HP, et al. Comparisons and impacts of the basic components of sarcopenia definition and their pairwise

combinations in gastric cancer: a large-scale study in a Chinese population. Front Nutr. 2021; 8: 709211. doi: 10.3389/fnut.2021.709211.
36. Huang DD, Yu DY, Wang WB, et al. Global leadership initiative in malnutrition (GLIM) criteria using hand-grip strength adequately predicts postoperative complications and long-term survival in patients underwent radical gastrectomy for gastric cancer. Eur J Clin Nutr. 2022; 76(9): 1323-1331. doi: 10.1038/s41430-022-01109-2.
37. Wang SL, Chen CB, Huang YS, et al. Muscle-specific strength is an alternative to muscle mass and grip strength for predicting outcomes in patients with gastric cancer. Eur J Surg Oncol. 2025; 51(9): 110229. doi: 10.1016/j.ejso.2025.110229.
38. Ding P, Wu H, Li T, et al. Impact of preoperative sarcopenia on postoperative complications and prognosis in patients undergoing robotic gastric cancer surgery: a propensity score matching study. Nutrition. 2024; 123: 112408. doi: 10.1016/j.nut.2024.112408.
39. Yamagishi S, Okamura Y, Kang W, et al. Impact of sarcopenic obesity on severe postoperative complications in patients with gastric cancer undergoing gastrectomy. Dig Surg. 2023; 40(5): 143-152. doi: 10.1159/000531797.
40. Juez LD, Priego P, Bajawi M, et al. Impact of sarcopenic obesity on long-term cancer outcomes and postoperative complications after gastrectomy for gastric cancer. J Gastrointest Surg. 2023; 27(1): 35-46. doi: 10.1007/s11605-022-05492-w.
41. Liu Q, Meng C, Cao S, et al. The whole-course of supervised home-based multi-modal prehabilitation to improve clinical outcome in patients undergoing neoadjuvant chemotherapy prior to gastrectomy: a single-center randomized controlled trial. Gastric Cancer. 2026; 29(3): 669-679. doi: 10.1007/s10120-026-01735-w.
42. Carvalho ALM, Gonzalez MC, Sousa IM, et al. Low skeletal muscle radiodensity is the best predictor for short-term major surgical complications in gastrointestinal surgical cancer: a cohort study. PLoS One. 2021; 16(2): e0247322. doi: 10.1371/journal.pone.0247322.
43. Cao Y, Heng Y, Song M, et al. Preoperative myosteatosis and intermuscular adiposity as CT-derived nutritional prognostic markers in colorectal cancer: a multicenter development-validation study. Clin Nutr. 2025; 53: 8-25. doi: 10.1016/j.clnu.2025.08.007.
44. Zhuang CL, Wu HF, Jiang HJ, et al. Muscle attenuation, not skeletal muscle index, is an independent prognostic factor for survival in gastric cancer patients with overweight and obesity. Nutrition. 2024; 122: 112391. doi: 10.1016/j.nut.2024.112391.
45. Tankel J, Wu W, Gold M, et al. Persistent and developed radiological myosteatosis, not sarcopenia, during neoadjuvant chemotherapy is associated with postoperative complications and survival among patients with locally advanced esophageal adenocarcinoma. Clin Nutr ESPEN. 2026; 73: 102975. doi: 10.1016/j.clnesp.2026.102975.
46. Tantisarasart T, Rachapongthai N, Tanasansuttiporn J, et al. Association of preoperative handgrip strength and postoperative recovery with outcomes in cardiac surgery patients ≥60 years old. JTCVS Open. 2026; 30: 101612. doi: 10.1016/j.xjon.2026.101612.
47. Li C, Chen Y, Wu H, et al. Influence of handgrip strength on postoperative complications and survival in primary liver cancer patients. Nutr Hosp. 2025; 42(2): 302-310. doi: 10.20960/nh.05564.
48. Rodrigues HHNP, Araujo KTP, Aguilar-Nascimento JE, et al. The 30-day readmission rate of patients with an overlap of probable sarcopenia and malnutrition undergoing major oncological surgery. Einstein (Sao Paulo). 2024; 22: eAO0733. doi: 10.31744/einstein_journal/2024AO0733.
49. Dong QT, Cai HY, Zhang Z, et al. Influence of body composition, muscle strength, and physical performance on the postoperative complications and survival after radical gastrectomy for gastric cancer: a comprehensive analysis from a large-scale prospective study. Clin Nutr. 2021; 40(5): 3360-3369. doi: 10.1016/j.clnu.2020.11.007.
50. Wu WY, Dong JJ, Huang XC, et al. AWGS2019 vs EWGSOP2 for diagnosing sarcopenia to predict long-term prognosis in Chinese patients with gastric cancer after radical gastrectomy. World J Clin Cases. 2021; 9(18): 4668-4680. doi: 10.12998/wjcc.v9.i18.4668.
51. Hofmann FO, Heiliger C, Tschaidse T, et al. Validation of body composition parameters extracted via deep learning-based segmentation from routine computed tomographies. Sci Rep. 2025; 15(1): 11909. doi: 10.1038/s41598-025-96238-6.
52. Smit KC, Derksen JWG, Kurk SA, et al. Use of automated assessment for determining associations of low muscle mass and muscle loss with overall survival in patients with colorectal cancer - a validation study. Clin Nutr ESPEN. 2024; 63: 572-584. doi: 10.1016/j.clnesp.2024.07.001.
53. Sun Y, Tian Y, Cao S, et al. Supervised multimodal prehabilitation and clinical outcomes in older patients with frailty and gastric cancer: the GISSG+2201 randomized clinical trial. JAMA Surg. 2026; 161(3): 223-233. doi: 10.1001/jamasurg.2025.6256.
54. Yamamoto K, Shinno N, Omori T, et al. Exercise-nutrition prehabilitation attenuates lean body mass loss before gastrectomy: a randomized controlled trial. Gastric Cancer. 2026; 29(4): 854-864. doi: 10.1007/s10120-026-01747-6.
55. Chen J, Hong C, Chen R, et al. Prognostic impact of a 3-week multimodal prehabilitation program on frail elderly patients undergoing elective gastric cancer surgery: a randomized trial. BMC Gastroenterol. 2024; 24(1): 403. doi: 10.1186/s12876-024-03490-7.
56. Yu D, Chen W, Zhang F, et al. Laparoscopic radical gastrectomy is a protective factor for clinical outcomes in gastric cancer patients with sarcopenia. Eur J Surg Oncol. 2025; 51(10): 110299. doi: 10.1016/j.ejso.2025.110299.
57. Jeong JY, Hwang J, Park SH, et al. Simple clinical parameters to identify sarcopenia 1 year after gastrectomy for gastric cancer. Gastric Cancer. 2026; 29(4): 841-853. doi: 10.1007/s10120-026-01748-5.
58. Zhao H, Dong Q, Chen C, et al. Perioperative body composition changes and their clinical implications in patients with gastric cancer undergoing radical gastric cancer surgery: a prospective cohort study. J Gastrointest Surg. 2025; 29(1): 101877. doi: 10.1016/j.gassur.2024.101877.