

# Percentile-based averaging and skeletal muscle gauge improve body composition analysis: validation at multiple vertebral levels

J. Peter Marquardt<sup>1,4\*</sup> , Eric J. Roeland<sup>2</sup>, Emily E. Van Seventer<sup>3</sup>, Till D. Best<sup>4,5</sup>, Nora K. Horick<sup>3</sup>, Ryan D. Nipp<sup>3</sup> & Florian J. Fintelmann<sup>4</sup> 

<sup>1</sup>Department of Radiology, RWTH Aachen University, Aachen, Germany; <sup>2</sup>Knight Cancer Institute, Oregon Health and Sciences University, Portland, OR, USA; <sup>3</sup>Department of Medicine, Division of Hematology and Oncology, Massachusetts General Hospital Cancer Center and Harvard Medical School, Boston, MA, USA; <sup>4</sup>Department of Radiology, Division of Thoracic Imaging and Intervention, Massachusetts General Hospital, Boston, MA, USA, and <sup>5</sup>Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Radiology, Berlin, Germany

## Abstract

**Background** Skeletal muscle metrics on computed tomography (CT) correlate with clinical and patient-reported outcomes. We hypothesize that aggregating skeletal muscle measurements from multiple vertebral levels and skeletal muscle gauge (SMG) better predict outcomes than skeletal muscle radioattenuation (SMRA) or -index (SMI) at a single vertebral level.

**Methods** We performed a secondary analysis of prospectively collected clinical (overall survival, hospital readmission, time to unplanned hospital readmission or death, and readmission or death within 90 days) and patient-reported outcomes (physical and psychological symptom burden captured as Edmonton Symptom Assessment Scale and Patient Health Questionnaire) of patients with advanced cancer who experienced an unplanned admission to Massachusetts General Hospital from 2014 to 2016. First, we assessed the correlation of skeletal muscle cross-sectional area, SMRA, SMI, and SMG at one or more of the following thoracic (T) or lumbar (L) vertebral levels: T5, T8, T10, and L3 on CT scans obtained  $\leq 50$  days before index assessment. Second, we aggregated measurements across all available vertebral levels using percentile-based averaging (PBA) to create the average percentile. Third, we constructed one regression model adjusted for age, sex, sociodemographic factors, cancer type, body mass index, and intravenous contrast for each combination of (i) vertebral level and average percentile, (ii) muscle metrics (SMRA, SMI, & SMG), and (iii) clinical and patient-reported outcomes. Fourth, we compared the performance of vertebral levels and muscle metrics by ranking otherwise identical models by concordance statistic, number of included patients, coefficient of determination, and significance of muscle metric.

**Results** We included 846 patients (mean age:  $63.5 \pm 12.9$  years, 50.5% males) with advanced cancer [predominantly gastrointestinal (32.9%) or lung (18.9%)]. The correlation of muscle measurements between vertebral levels ranged from 0.71 to 0.84 for SMRA and 0.67 to 0.81 for SMI. The correlation of individual levels with the average percentile was 0.90–0.93 for SMRA and 0.86–0.92 for SMI. The inpatient correlation of SMRA with SMI was 0.21–0.40. PBA allowed for inclusion of 8–47% more patients than any single-level analysis. PBA outperformed single-level analyses across all comparisons with average ranks 2.6, 2.9, and 1.6 for concordance statistic, coefficient of determination, and significance (range 1–5,  $\mu = 3$ ), respectively. On average, SMG outperformed SMRA and SMI across outcomes and vertebral levels: the average rank of SMG was 1.4, 1.4, and 1.4 for concordance statistic, coefficient of determination, and significance (range 1–3,  $\mu = 2$ ), respectively.

**Conclusions** Multivertebral level skeletal muscle analyses using PBA and SMG independently and additively outperform analyses using individual levels and SMRA or SMI.

**Keywords** Body composition analysis; Skeletal muscle; Patient-reported outcomes; Sarcopenia; Survival; Skeletal muscle; BMI; Skeletal muscle gauge

Received: 13 February 2021; Revised: 27 September 2021; Accepted: 5 October 2021

\*Correspondence to: J. Peter Marquardt, Department of Radiology, RWTH Aachen University, Aachen, Pauwelsstraße 30, 52074 Aachen, Germany. Email: peter.marquardt@rwth-aachen.de

J. Peter Marquardt and Eric J. Roeland shared first authorship.

Ryan D. Nipp and Florian J. Fintelmann shared last authorship.

## Introduction

Clinical studies increasingly utilize body composition analysis, particularly skeletal muscle assessment, to refine risk stratification of patients with cancer.<sup>1–3</sup> Segmentation of muscle and adipose tissue on computed tomography (CT) at the level of the third lumbar vertebra (L3) represents the current *de facto* gold standard method for opportunistic body composition analysis on CT scans obtained for routine clinical care.<sup>2,4–6</sup> Skeletal muscle cross-sectional area (CSA) at the L3 level correlates better with whole-body muscle mass than CSA at other vertebral levels, thereby rendering L3 the most common measurement location for body composition analysis.<sup>7–10</sup> Muscle CSA can be divided by the squared height (in meters) to obtain the skeletal muscle index (SMI). SMI can be multiplied with skeletal muscle radioattenuation (SMRA) to calculate skeletal muscle gauge (SMG). While prior research suggests that SMG is superior to CSA and attenuation alone, comparisons with other body composition metrics (e.g. SMI) are limited.<sup>3,11–17</sup>

Segmentation at multiple vertebral levels provides a better assessment of an individual's muscle and adipose tissue<sup>9</sup> and may allow for body composition analysis in patients lacking CT imaging at the L3 level. However, single-level analyses remain the default strategy, in part because of the time required for manual segmentation, and few studies have aggregated muscle measurements across multiple vertebral levels to predict clinical outcomes.<sup>18</sup> Recent advances in machine learning have fuelled the creation of algorithms that can replace manual segmentation at various vertebral levels, thereby enabling large-scale, multilevel body composition analysis.<sup>19–22</sup> These algorithms can provide radioattenuation in addition to CSA, thereby making SMG widely available.

Multivertebral level body composition analysis introduces a new set of challenges concerning data analysis and interpretation. For example, the high correlation of skeletal muscle measurements between vertebral levels<sup>7,11,17,23,24</sup> would lead to collinearity if measurements were included as separate variables within the same statistical model. Although aggregation of measurements from multiple vertebral levels into a single metric would prevent collinearity during statistical analysis, established methods for aggregation and interpretation of the aggregate are currently lacking. Also, patients might be missing measurements at one or more vertebral level, and methods to address missing values in multivertebral level analyses are needed.

The objectives of this paper are threefold: (i) describe and discuss the concept of percentile-based averaging, a novel, flexible, and scalable method for aggregating multivertebral level body composition measurements into a single metric while also addressing missing values; (ii) validate percentile-based averaging in a cohort of patients with advanced cancer by comparing it with conventional complete-case single-level-analyses; and (iii) substantiate evidence supporting the use of SMG.

We hypothesize that aggregating skeletal muscle measurements from multiple vertebral levels using percentile-based averaging and SMG better predict clinical and patient-reported outcomes than SMRA or SMI at a single vertebral level.

An improved understanding of multivertebral level body composition analysis would represent a significant advance for cancer cachexia and sarcopenia research.

## Methods and materials

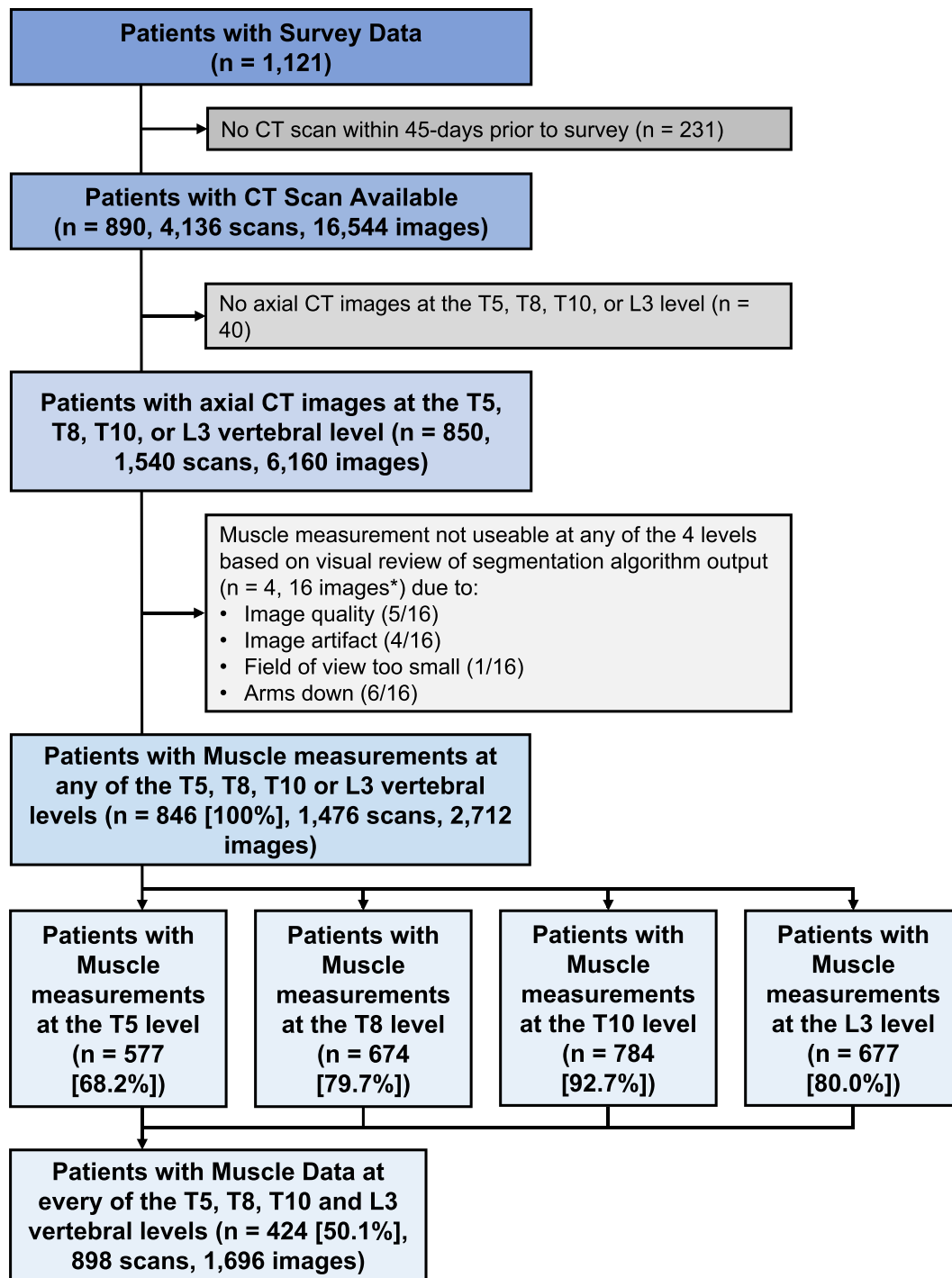
This is a secondary analysis of a prospective longitudinal cohort study conducted between September 2014 and May 2016 with approval from the Dana Farber/Harvard Cancer Center institutional review board.<sup>25</sup> In brief, we recruited adult patients with advanced cancer who experienced an unplanned admission to Massachusetts General Hospital to describe their clinical and patient-reported outcomes.<sup>25</sup>

### Study design and patient cohort

In the original study, we approached patients within 5 days after their first unplanned hospital admission and collected patient-reported outcomes following enrolment using standardized surveys.<sup>26–28</sup> We also analysed CT scans obtained within 50 days prior to index assessment, as previously described.<sup>29</sup> We included patients with CT scans that technically allowed evaluation of muscle at one or more of the following thoracic (T) or lumbar (L) vertebral levels: T5, T8, T10, and L3 (*Figure 1*).

### Skeletal muscle and subcutaneous adipose tissue measurements

We used a multistep process (pipeline) including previously validated machine learning algorithms to (i) identify vertebral

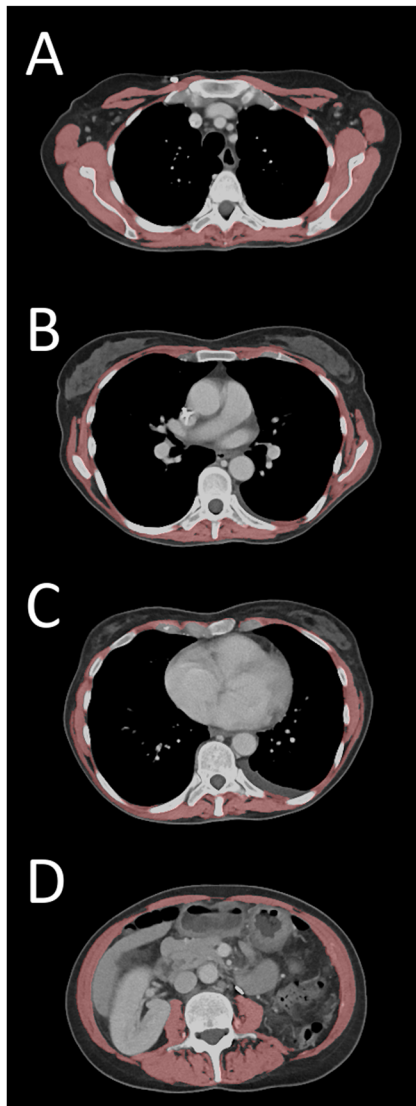


**Figure 1** Flow diagram of inclusion and exclusion criteria. Abbreviations: CT, computed tomography; T5/T8/T10, fifth/eighth/tenth thoracic; L3, third lumbar. \*The last exclusion step excludes patients who had unusable segmentations at all four vertebral levels. We determined segmentation quality for each image individually, which is why we report errors for each of the four slices.

levels T5, T8, T10, and L3 on axial images and (ii) segment skeletal muscle at each level using thresholds of  $-29$  to  $+150$  Hounsfield Units.<sup>20</sup>

The output consisted of CSA and mean attenuation of all voxels labelled as muscles, as well as a label map (Figure 2).

We subjected each label map to independent review by two trained analysts (medical students with 1 year of experience with body composition analysis, trained and supervised by a board-certified radiologist) blinded to clinical and patient-reported outcomes. Analysts determined the pres-



**Figure 2** Illustration skeletal muscle segmentation (red) at the level of the fifth (A), eighth (B), and tenth (C) thoracic, and third lumbar (D) vertebral body in a 65-year-old female with advanced pancreatic cancer.

ence or absence of intravenous contrast and discarded instances where the labels did not match anatomy, as previously described.<sup>29</sup> We calculated SMI [ $\text{cm}^2/\text{m}^2$ ] by dividing CSA of skeletal muscle by the patient's squared height. We calculated SMG [ $\text{Hounsfield Unit}\cdot\text{cm}^2/\text{m}^2$ ] by multiplying SMI with SMRA.

### Percentile-based averaging

We used percentile-based averaging to aggregate body composition measurements from multiple vertebral levels into a single, normalized value: the average percentile. For each body composition metric (SMRA, SMI, & SMG), and at every analysed vertebral level, we ranked patients within a stratum.

Strata were defined by characteristics known to influence body composition measurements<sup>30,31</sup>. We stratified SMI by sex, and stratified SMRA and SMG by sex and presence of intravenous contrast. We then normalized the rank to a percentile by multiplying the rank with 100 and dividing the product by the number of cases in the stratum. Then, we calculated the average percentile of each metric for each patient by averaging the percentiles at each vertebral level. We did not include missing values in calculating the average percentile, which is mathematically equivalent to first imputing missing percentiles using the mean value of non-missing levels (Appendix 1, Proof 1). Therefore, we assigned an average percentile to every patient with one or more measurements for a given metric. For example, a female patient's SMI at the L3 level would be ranked against the L3 SMI of other female patients to obtain her L3 SMI percentile. This percentile would then be averaged with her T8 and T10 SMI percentiles, while a missing value at T5 would not contribute to her average percentile.

We implemented percentile-based averaging in R version 4.0.2 (Vienna, Austria).<sup>32</sup> The source code of the implementation is available at <https://github.com/p-mq/Percentiles>, and the packaged version may be downloaded at the Comprehensive R Archive Network as 'percentiles' (<https://cran.r-project.org/web/packages>).

### Clinical and patient-reported outcomes

We used prospectively collected clinical and patient-reported outcomes, as previously described.<sup>25</sup> Briefly, clinical outcomes included overall survival, hospital length of stay, time to unplanned hospital readmission or death, and readmission or death within 90 days. Patient-reported outcomes included physical and psychological symptom burden captured as the revised Edmonton Symptom Assessment System total (range 0–100) and physical score (range 0–70)<sup>27,28</sup> as well as the Patient Health Questionnaire-4 consisting of depression and anxiety scores (both range 0–6).<sup>26</sup> Higher scores indicate greater symptom burden across all patient-reported outcomes.

### Statistical modelling and model comparison

We performed analyses and visualizations in R version 4.0.2 (Vienna, Austria) using percentiles, BlanketStatsments, survAUC, ggplot2, and psych libraries<sup>32</sup> and Python version 3.8. Our code for this analysis is available for download at [https://github.com/p-mq/Percentile\\_based\\_averaging](https://github.com/p-mq/Percentile_based_averaging).

We analysed the relationship between muscle metrics and measurement locations using Pearson's correlation coefficient. We used multivariable Cox regression to model the relationship between muscle metrics and overall survival and

time to hospital readmission or death. We used multivariable logistic regression to model the relationship between muscle metrics and readmission or death within 90 days. We used multivariable linear regression to model the relationship between muscle metrics and hospital length of stay as well as patient-reported outcomes. We adjusted all models for potential confounders: age, sex, marital status, education level, insurance, cancer type, and body mass index (BMI).<sup>29</sup> If the muscle metric was SMRA or SMG, we additionally adjusted for presence of intravenous contrast.<sup>30</sup>

We evaluated a total of 240 statistical models. Of these, 120 models reflected all combinations of the three different muscle metrics (SMRA, SMI, & SMG), five measurement locations (T5, T8, T10, L3, and average percentile), and eight outcomes described above. For internal validation and exploration, we generated an analogous set of 120 models using the subset of patients with muscle measurements at all four vertebral levels.

We ranked statistical models describing the same outcome according to four pre-specified performance characteristics: (i) concordance statistic (*C*-statistic) as a measure of in-group predictive accuracy was the primary performance characteristic; (ii) number of included patients (*n*); (iii) coefficient of determination ( $R^2$ ); and (iv) significance of muscle metric (*P* value). We considered larger values to represent better performance for *n*, *C*-statistic, and  $R^2$ , while smaller values indicated better performance for *P* values. We used Uno's modified *C*-statistic for right-censored outcomes in Cox regression models.<sup>33</sup>

For each of the four performance characteristics, we assigned a rank for measurement location and a rank for muscle metric to each model. Ranks ranged from 1 to 3 for muscle metrics and from 1 to 5 for measurement location. We assigned ranks in descending order. We assigned the same, best available rank to tied models. We calculated the average rank of performance characteristics for each measurement location and each muscle metric by taking the mean of individual model ranks using the respective measurement location or metric. If there were no differences between measurement locations and muscle metrics, all models' average ranks would be equal to the expected value ( $\mu$ ). The average rank of *C*-statistic was the key metric for the primary and secondary hypothesis. Construction and comparison of statistical models was implemented in R version 4.0.2 (Vienna, Austria).<sup>32</sup> The source code of the implementation is available at <https://github.com/p-mq/BlanketStatsments> and the packaged version may be downloaded at the Comprehensive R Archive Network as 'BlanketStatsments' (<https://cran.r-project.org/package=BlanketStatsments>).

We tested for differences in demographic variables, cancer history, Charlson comorbidity index, and BMI between patients with and without imaging. We used *t*-tests for continuous variables, Fisher's exact test for race/ethnicity, and  $\chi^2$  tests for all other categorical variables.

## Results

### Patient sample

Of 1121 patients enrolled in the original study, 890 (79.4%) had one or more CT scans within 50 days prior to the index assessment. We excluded 44 (4.9%) patients who lacked measurements for all vertebral levels of interest due to lack of available images or artefacts precluding accurate muscle segmentation (Figure 1). The remaining 846 participants (1476 scans imaging 2712 levels that passed quality assurance) had a mean age of  $63.5 \pm 12.9$  years, and 50.5% were male participants (Table 1). Participants were predominantly white (92.1%), married (65.8%), and educated beyond high school (59.7%). The most prevalent cancer types were gastrointestinal (32.9%), lung (18.9%), and genitourinary (9.0%), with a mean time since diagnosis of  $3.2 \pm 4.3$  years. The mean BMI was  $25.9 \pm 5.96$  kg/m<sup>2</sup>, and most (82.8%) analysed CT images were performed with intravenous contrast. The median time to death or loss to follow-up was 117 days with an interquartile range of 214 days. The mortality and readmission rate within 90 days was 42.0% and 44.3%, respectively, with a mean hospital length of stay of  $6.7 \pm 5.6$  days. Patients excluded due to lack of available imaging were older (2.5 years), more likely to have government-sponsored health insurance, and more likely to carry a diagnosis of advanced cancer for a longer period of time (all  $P < 0.01$ ) (Table A1). No differences were found for sex, race/ethnicity, marital status, education level, Charlson comorbidity index score, and BMI (Table A1).

### Skeletal muscle measurements

Muscle measurements were available at 4, 3, 2, and only 1 of 4 possible vertebral levels in 424 (50%), 228 (27%), 138 (16%), and 56 (7%) of 846 (100%) patients. Muscle measurements were most frequently available at the T10 level (93%) and least frequently available at the T5 level (68%) (Figure 1). At L3, muscle measurements were available for 80.0% of patients.

### Correlation of muscle metrics between vertebral levels

The correlation of muscle metrics between vertebral levels was high for SMRA (range 0.71–0.84) and SMI (range 0.67–0.81). The average percentile demonstrated the best correlation with each assessed level (range 0.90–0.93 for SMRA and 0.86–0.92 for SMI) (Figure 3). SMRA only correlated moderately with SMI (range 0.10–0.40) (Figure 3).



**Table 1** Patient characteristics

Characteristic	Overall cohort	
	<i>N</i> = 846	%
Age—mean (SD)	63.46 (12.90)	
Sex		
Male	427	50.5
Female	419	49.5
Race/ethnicity		
White	779	92.1
African American	30	3.5
Asian	20	2.4
Hispanic	15	1.8
Other	2	0.1
Marital status		
Married	557	65.8
Single	135	16.0
Divorced	92	10.9
Widowed	62	7.3
Education level		
High school and below	266	31.4
Beyond high school	505	59.7
Declined to provide	75	8.9
Health insurance		
Government-sponsored	421	49.8
Private	421	49.8
None	4	0.5
Cancer type		
Gastrointestinal	278	32.9
Lung	160	18.9
Genitourinary	76	9.0
Melanoma	68	8.0
Breast	56	6.6
Sarcoma	45	5.3
Gynaecologic	42	5.0
Head and neck	41	4.8
Leukaemia	35	4.1
Lymphoma	33	3.9
Cancer of unknown primary	12	1.4
Months since cancer diagnosis—mean (SD)	38.44 (51.60)	
Charlson comorbidity index score—mean (SD)	0.84 (1.25)	
Body mass index—mean (SD)	25.87 (5.96)	
Measurements with IV contrast ( <i>N</i> = 2712 <sup>a</sup> )	2245 <sup>a</sup>	82.7 <sup>a</sup>

CT, computed tomography; IV, intravenous; SD, standard deviation.

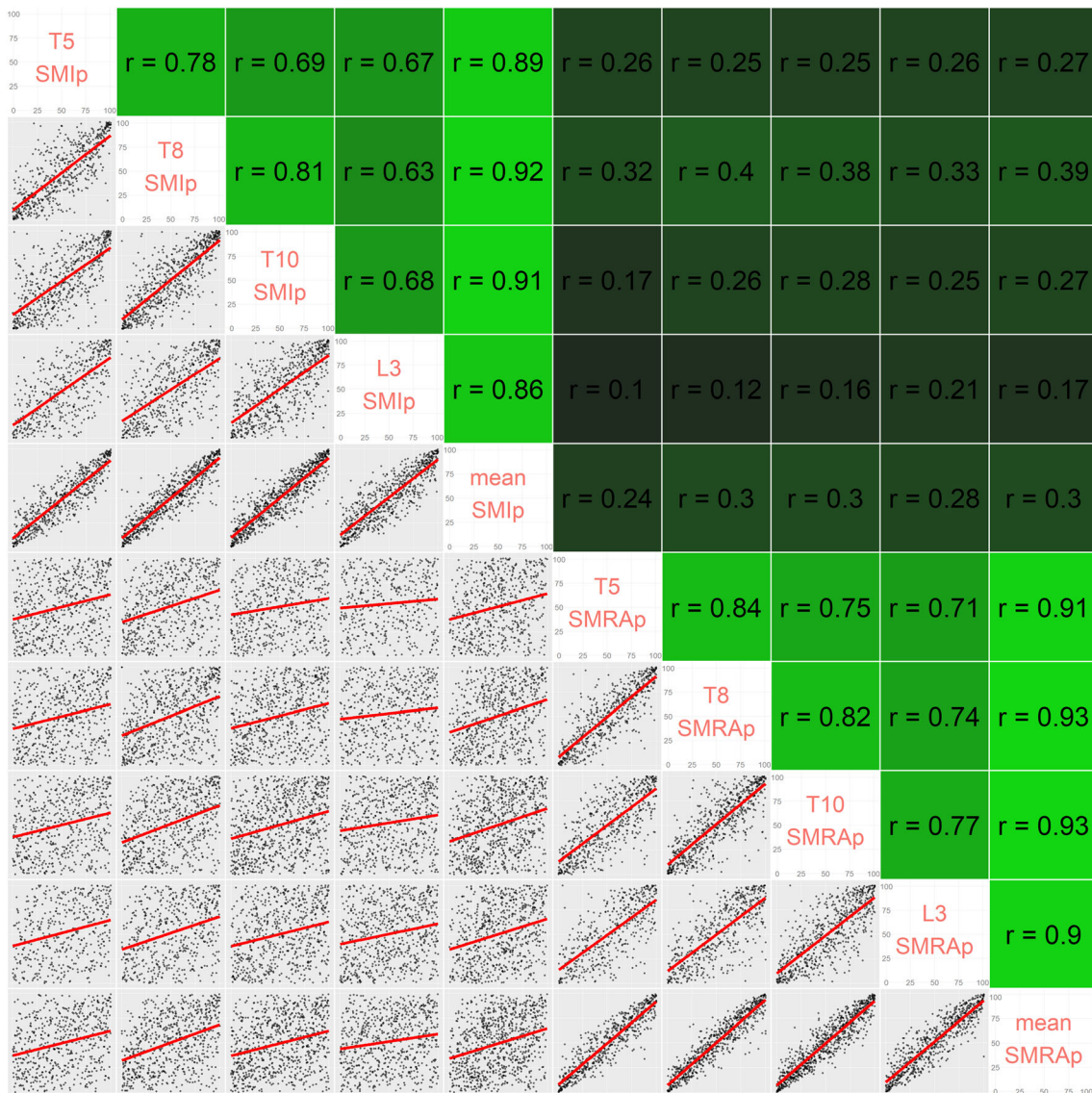
<sup>a</sup>The proportion of measurements with intravenous contrast present was calculated on a per slice basis.

### Performance of single-level complete-case analysis versus percentile-based averaging

We found that percentile-based averaging aggregating muscle measurements from any available vertebral level performed better than complete-case analyses at the four possible single vertebral levels for all four chosen performance metrics: in otherwise identical models compared by ranking on a scale from 1 to 5 (expected value of 3), models using percentile-based averaging had an average *C*-statistic rank of 2.6 compared with a mean of 3.1 (range 2.3–3.8) in single-level models. In all models, percentile-based averaging maximized the number of included patients (average rank 1.0). The increase in the number of included patients achieved by percentile-based averaging was 25% compared with complete-case analyses at the L3 level. The average rank for *R*<sup>2</sup> value (2.9) and *P* value (1.6) was also better for

percentile-based averaging than for single-level analyses, indicating superior performance for percentile-based averaging than expected under the null hypothesis of equal performance (Table 2, Figure 4).

In the subset of 424 patients with muscle measurements at all four levels, we found that percentile-based averaging demonstrated superior performance compared with single-level analyses: in otherwise identical models compared by ranking on a scale from 1 to 5 (expected value of 3), the average rank for *C*-statistics was 2.2 for models with percentile-based averaging, compared with a mean of 3.2 (range 3.0–3.5) for single-level models. Similarly, mean ranks for *R*<sup>2</sup> values [2.39 vs. 3.2 (range 2.6–3.8)] and *P* values [2.1 vs. 3.2 (range 2.8–3.8)] were also lower for percentile-based averaging, indicating superior performance for percentile-based averaging than expected under the null hypothesis of equal performance (Table 2, Figure 4).



**Figure 3** Correlation of skeletal muscle index and skeletal muscle radioattenuation across the T5, T8, T10, and L3 levels. Bottom left half: dot plots and regression estimates (red line); top right half: Pearson's correlation coefficients, brighter green corresponding to higher correlation. Top-left quadrant: correlation of skeletal muscle index percentiles at each vertebral level. Bottom right quadrant: correlation of skeletal muscle radioattenuation percentiles at each vertebral level. Top right and bottom left quadrant: correlation of skeletal muscle index percentiles and skeletal muscle radioattenuation percentiles at each vertebral level. Abbreviations: L3, third lumbar; SMIp, skeletal muscle index percentile; SMRAp, skeletal muscle radioattenuation percentile; T5/T8/T10, fifth/eight/tenth thoracic.

We chose to not to highlight associations of measurement location with outcomes as these associations do not contribute to evaluating the proposed changes in methodology and would require extensive corrections for multiple testing. For results of individual models, refer to *Table A2*.

### Comparison of muscle metrics

Models based on SMG generally outperformed models based on SMRA and SMI with regards to *C*-statistics: In otherwise

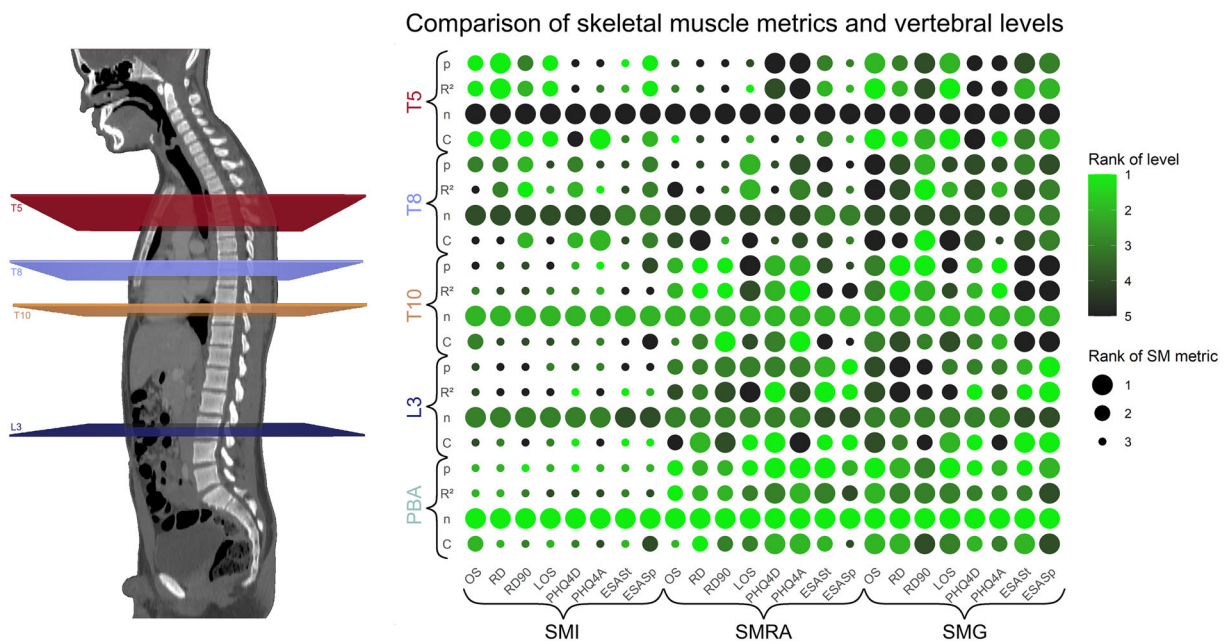
identical models compared by ranking on a scale from 1 to 3 (expected value of 2), the average rank of *C*-statistic was 1.4 for SMG-based models, indicating superior performance compared with SMRA-based and SMI-based models. Average ranks of  $R^2$  and *P* values of SMG-based models were 1.4 and 1.4, indicating superior performance compared with SMRA-based and SMI-based models (*Table 3, Figure 4*).

We found the same trend in the subset of 424 patients with muscle measurements available at all levels, with mean ranks of SMG-based models being 1.6, 1.3, and 1.3 for *C*-statistic,  $R^2$ , and *P* value, respectively (*Table 3, Figure 4*).

**Table 2** Performance of vertebral levels T5, T8, T10, and L3 or multilevel analysis using percentile-based averaging compared against each other in concordance statistic, number of included patients, coefficient of determination, and *P* value using otherwise equal statistical models

Model specifications		Average rank of model performance metrics ( $\mu = 3$ )			
Measurement location	Included patients	<i>C</i>	<i>N</i>	<i>R</i> <sup>2</sup>	<i>P</i>
T5	Measurements at any level ( <i>n</i> = 846)	2.33	5.00	2.67	3.33
	Measurements at all levels ( <i>n</i> = 424)	3.25	—	3.46	3.33
T8	Measurements at any level ( <i>n</i> = 846)	3.75	3.75	3.17	3.71
	Measurements at all levels ( <i>n</i> = 424)	3.46	—	3.83	3.83
T10	Measurements at any level ( <i>n</i> = 846)	3.58	2.00	3.04	3.00
	Measurements at all levels ( <i>n</i> = 424)	3.13	—	2.58	2.75
L3	Measurements at any level ( <i>n</i> = 846)	2.71	3.25	3.21	3.38
	Measurements at all levels ( <i>n</i> = 424)	3.00	—	2.83	3.00
Percentile-based averaging	Measurements at any level ( <i>n</i> = 846)	2.63	1.00	2.92	1.58
	Measurements at all levels ( <i>n</i> = 424)	2.17	—	2.29	2.08

$\mu$  describes the expected ranking if there was no difference between methods. *C*, concordance statistic; *N*, number of included patients; *R*<sup>2</sup>, coefficient of determination; *P*, *P* value;  $\mu$  = expected value.



**Figure 4** Comparison of skeletal muscle metrics (*x*-axis/dot size) and vertebral levels (*y*-axis/colour) across concordance statistic, number of included patients, coefficient of determination, and *P* value. Abbreviations: *C*, concordance statistic; ESASp, Edmonton Symptom Assessment System physical; ESASSt, Edmonton Symptom Assessment System total; L3, third lumbar; LOS, hospital length of stay; *n*, number of included patients; OS, overall survival; *P*, *P* value; PBA, percentile-based averaging; PHQ4A, patient health Questionnaire-4 anxiety; PHQ4D, patient health Questionnaire-4 depression; *R*<sup>2</sup>, coefficient of determination; RD, time to unplanned hospital readmissions or death; RD90, readmission or death within 90 days; SM, skeletal muscle; SMG, skeletal muscle gauge; SMI, skeletal muscle index; SMRA, skeletal muscle radioattenuation; T5/T8/T10, fifth/eighth/tenth thoracic.

**Table 3** Performance of skeletal muscle metrics compared against each other in concordance statistic, number of included patients, coefficient of determination, and *P* value using otherwise equal statistical models

Model specifications		Average rank of model performance metrics ( $\mu = 2$ )			
Metric	Included patients	<i>C</i>	<i>N</i>	<i>R</i> <sup>2</sup>	<i>P</i>
Skeletal muscle radioattenuation	Measurements at any level ( <i>n</i> = 846)	2.10	1.00	1.85	1.90
	Measurements at all levels ( <i>n</i> = 424)	2.08	1.00	2.18	2.23
Skeletal muscle index	Measurements at any level ( <i>n</i> = 846)	2.55	1.00	2.75	2.70
	Measurements at all levels ( <i>n</i> = 424)	2.35	1.00	2.58	2.50
Skeletal muscle gauge	Measurements at any level ( <i>n</i> = 846)	1.35	1.00	1.40	1.40
	Measurements at all levels ( <i>n</i> = 424)	1.58	1.00	1.25	1.28

*C*, concordance statistic; *N*, number of included patients; *R*<sup>2</sup>, coefficient of determination; *P*, *P* value;  $\mu$  = expected value.



We chose to not to highlight associations of muscle metrics with outcomes as these associations do not contribute to evaluating the proposed changes in methodology and would require extensive corrections for multiple testing. For results of individual models, refer to *Table A2*.

## Discussion

In this study, we describe and validate a novel approach to harness multivertebral level muscle measurements using percentile-based averaging and validate SMG as a muscle metric superior to SMRA and SMI. We evaluated percentile-based averaging and SMG in a large cohort of patients with advanced cancer using prospectively collected clinical and patient-reported outcomes. We found that percentile-based averaging of multivertebral level muscle measurements and SMG additively and independently outperformed the standard complete-case, single-level approach using SMI at the L3 level for outcome prediction and allowed for the inclusion of more patients. Collectively, these findings indicate that percentile-based averaging of multivertebral level body composition measurements and SMG improve body composition analysis.

Skeletal muscle gauge combines aspects of both muscle quantity and quality. We demonstrated that SMG better predicted clinical and patient-reported outcomes compared with SMRA or SMI alone.<sup>14–16</sup> SMI is the most widely used muscle metric but only measures muscle quantity.<sup>4,5</sup> SMRA is weakly correlated with SMI across vertebral levels, suggesting that these metrics capture distinct aspects. Future studies should consider SMG as the metric encompassing both muscle quantity and quality and establish reference values for SMG.

There are multiple reasons to use percentile-based averaging in addition to SMG. One, while multilevel body composition measurements are increasingly available, methods to aggregate this information are missing.<sup>18,24,34,35</sup> Two, the current study's performance characteristics indicate that multilevel body composition analysis using percentile-based averaging can improve upon the current standard of single-level complete-case analysis. We chose the *C*-statistic as a measure of predictive accuracy for the primary characteristic, as the *C*-statistic reflects a model's primary purpose. *C*-statistics were generally higher for models using percentile-based averaging, suggesting multilevel analysis improves prediction and prognostication of clinical outcomes and patient-reported outcomes. Three, multilevel analysis increased the number of included patients compared with single-level analyses and reduced standard errors, aligned with previous work.<sup>9,10</sup> Four, the associated increased statistical power may allow investigators to minimize sample size while also minimizing their exclusion rate. Minimizing the exclusion rate would address non-random missingness

due to incompletely imaged vertebral levels or artefacts, both of which frequently limit body composition studies relying on CT scans acquired as part of routine clinical care.<sup>3</sup>

Percentile-based averaging has additional advantages. One, clinicians frequently use percentiles, and thus percentile-based averaging represents a familiar and approachable concept for clinicians. Two, percentile-based averaging is level-agnostic, meaning it enables investigators to assess measurements at vertebral levels without established reference values. Three, percentile-based averaging is scalable and can accommodate measurements from any number of vertebral levels. While each included vertebral level contributes equally to the average percentile, the interpretation is independent of the number or choice of vertebral levels. Specifically, the range of values remains 0–100 regardless of whether measurements originated from only the third lumbar or all thoracic and lumbar vertebral levels, and the differences in absolute CSA between measurement locations are no longer important. Four, percentile-based averaging also does not require dedicated statistical software and extensive statistical expertise, unlike other imputation methods.<sup>36</sup> We implemented code to use percentile-based averaging in R and made it freely available on the Comprehensive R Archive Network (<https://cran.r-project.org/package=percentiles>). Five, the average percentiles created with percentile-based averaging are fixed for a given dataset, contrary to other imputation methods.<sup>36</sup> As a result, investigators can store and transfer average percentiles between software programmes and collaborators. Six, the ability to estimate extreme values mitigates potential non-random missingness due to the exclusion of patients with incompletely imaged muscle due to limited field of view. In this study, patients excluded because of absent or unusable imaging did not have a significantly higher BMI than included patients, contrary to other reports. Seven, percentile-based averaging is flexible. While we focused on percentile-based averaging for muscle measurements in this study, the correlation of adipose tissue suggests that percentile-based averaging can also analyse multilevel adipose tissue measurements.<sup>37,38</sup>

Despite these advantages, percentile-based averaging is not universally applicable: For one, percentile-based averaging creates cohort-specific average percentiles. Intercohort comparison of these cohort-specific percentiles would require adjusting for intercohort differences in absolute values. Second, in small or unbalanced cohorts, stratification by sex and intravenous contrast in the process of calculating percentiles may result in small subgroups. If reference values were available for all analysed vertebral levels, level-specific percentiles of a given cohort could be compared with the reference cohort prior to averaging. Such standardized percentiles would resolve the limitations of intercohort comparisons and small subgroups described above. Unfortunately, we currently lack reference values for many vertebral levels

outside the abdomen, an important opportunity for future research.<sup>22</sup>

Our data presented does not allow us to conclusively answer the question what number of vertebral levels and which vertebral levels would be optimal. First, we found that no individual vertebral level performs above average across all performance characteristics (Table 2, Figure 4). However, the correlation between muscle measurements decreases as distance increases (Figure 3). Therefore, we assume the benefits of added levels to increase with anatomic distance. We further believe that the incremental benefit of each additional level decreases with the total number of levels. The optimal level for single-level analysis depends on the relationship of the outcome of interest with the muscles imaged at a given level and the characteristics of the patient cohort.<sup>39</sup>

We need to interpret these results in the context of several limitations. First, we conducted this study at a single, tertiary care institution in a population with limited demographic diversity. In contrast, the heterogeneity of our cohort regarding cancer type, treatments, and outcomes can introduce unaccounted variability but also suggests wider generalizability. Although we collected clinical and patient-reported outcomes prospectively, images were obtained as part of routine clinical care, and thus collected and analysed retrospectively. The compared statistical models only varied in one predictor, and therefore, we did not analyse individual combinations of multiple vertebral levels or muscle metrics. We compared and ranked the performance of different statistical models without testing for statistical significance of the performance differences. Given the multidimensionality of comparisons (measurement location, metric, outcome, & performance characteristic), reporting *P* values may not accurately encapsulate this study's complex and descriptive nature. We validated our findings internally using the data set with muscle data at all four levels, and we plan to validate the dataset externally in the future.

In conclusion, this study validates percentile-based averaging as a novel, flexible, and scalable method that permits aggregation of multiple vertebral levels of body composition data while also addressing missing values at one or more vertebral levels. Aggregating muscle measurements from multiple vertebral levels into a single metric

using percentile-based averaging improved predictive performance compared with complete-case, single-level analyses, including the current *de facto* gold standard using SMI at the L3 level. We further highlighted how SMG, as an indicator of both muscle quantity and quality, captures more information than muscle quantity or quality alone. Findings from this work should inform future body composition studies which stand to benefit from increased accuracy, statistical power, and interpretability conferred by this approach. We made the code to use percentile-based averaging freely available on the Comprehensive R Archive Network (<https://cran.r-project.org/package=percentiles>) to facilitate use and further validation.

## Acknowledgements

### Ethics declaration

This study is a secondary analysis of a study approved by the Dana Farber/Harvard Cancer Center institutional review board and performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.<sup>25,40</sup> All subjects gave their informed consent prior to their inclusion in the study.

## Conflicts of interest

F.J.F. was supported by the American Roentgen Ray Society Scholarship for this study and has a related patent pending. E.J.R., MD has served as a consultant for Mitobridge Inc., Asahi Kasei Pharmaceuticals, DRG Consulting, Napo Pharmaceuticals, American Imaging Management, Immuneering Corporation, and Prime Oncology. Additionally, he has served on recent advisory boards for Heron Pharmaceuticals, Vector Oncology, and Helsinn Pharmaceuticals. He has also served as a member on data safety monitoring boards for Oragenics, Inc, Galera Pharmaceuticals, and Enzychem Lifesciences Pharmaceutical Company. The other authors do not report relevant conflicts of interest.

## References

1. Martin L, Birdsell L, MacDonald N, Reiman T, Clandinin MT, McCargar LJ, et al. Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. *J Clin Oncol* 2013;**31**: 1539–1547.
2. Mourtzakis M, Prado CMM, Lieffers JR, Reiman T, McCargar LJ, Baracos VE. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Appl Physiol Nutr Metab* 2008;**33**: 997–1006.
3. Troschel AS, Troschel FM, Best TD, Gaijssert HA, Torriani M, Muniappan A, et al. Computed tomography-based body composition analysis and its role in lung cancer care. *J Thorac Imaging* 2020;**35**:91–100.
4. Chen L-K, Liu L-K, Woo J, Assantachai P, Auyeung T-W, Bahyah KS, et al. Sarcopenia in Asia: consensus report of the Asian Working Group for Sarcopenia. *J Am Med Dir Assoc* 2014;**15**:95–101.
5. Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyère O, Cederholm T, et al.

- Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing* 2019;**48**:16–31.
6. Pickhardt PJ, Graffy PM, Zea R, Lee SJ, Liu J, Sandfort V, et al. Automated CT biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study. *Lancet Digit Health* 2020;**2**:e192–e200.
  7. Derstine BA, Holcombe SA, Ross BE, Wang NC, Su GL, Wang SC. Skeletal muscle cutoff values for sarcopenia diagnosis using T10 to L5 measurements in a healthy US population. *Sci Rep* 2018;**8**:11369.
  8. Halpenny DF, Goncalves M, Schwitzer E, Golia Pernicka J, Jackson J, Gandelman S, et al. Computed tomography-derived assessments of regional muscle volume: validating their use as predictors of whole body muscle volume in cancer patients. *Br J Radiol* 2018;**91**:20180451.
  9. Shen W, Punyanyita M, Wang Z, Gallagher D, St-Onge M-P, Albu J, et al. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. *J Appl Physiol* 2004;**97**:2333–2338.
  10. Faron A, Luetkens JA, Schmeel FC, Kuetting DLR, Thomas D, Sprinkart AM. Quantification of fat and skeletal muscle tissue at abdominal computed tomography: associations between single-slice measurements and total compartment volumes. *Abdom Radiol* 2019;**44**:1907–1916.
  11. Derstine BA, Holcombe SA, Goulson RL, Ross BE, Wang NC, Sullivan JA, et al. Quantifying sarcopenia reference values using lumbar and thoracic muscle areas in a healthy population. *J Nutr Health Aging* 2018;**22**(1):180–185. <https://doi.org/10.1007/s12603-017-0983-3>
  12. Lee K, Shin Y, Huh J, Sung YS, Lee I-S, Yoon K-H, et al. Recent issues on body composition imaging for sarcopenia evaluation. *Korean J Radiol* 2019;**20**:205–217.
  13. Faron A, Sprinkart AM, Kuetting DLR, Feisst A, Isaak A, Enderl C, et al. Body composition analysis using CT and MRI: intra-individual intermodal comparison of muscle mass and myosteatosis. *Sci Rep* 2020;**10**:11765.
  14. Williams GR, Deal AM, Muss HB, Weinberg MS, Sanoff HK, Nyrop KA, et al. Skeletal muscle measures and physical function in older adults with cancer: sarcopenia or myopenia? *Oncotarget* 2017;**8**:33658–33665.
  15. Weinberg MS, Shachar SS, Muss HB, Deal AM, Popuri K, Yu H, et al. Beyond sarcopenia: characterization and integration of skeletal muscle quantity and radiodensity in a curable breast cancer population. *Breast J* 2018;**24**:278–284.
  16. Shachar SS, Deal AM, Weinberg M, Nyrop KA, Williams GR, Nishijima TF, et al. Skeletal muscle measures as predictors of toxicity, hospitalization, and survival in patients with metastatic breast cancer receiving taxane-based chemotherapy. *Clin Cancer Res* 2017;**23**:658–665.
  17. Mishra A, Bigam KD, Extermann M, Faramand R, Thomas K, Pidala JA, et al. Sarcopenia and low muscle radiodensity associate with impaired FEV1 in allogeneic haematopoietic stem cell transplant recipients. *J Cachexia Sarcopenia Muscle* 2020;**11**:1570–1579.
  18. Best TDCM, Mercaldo SF, Bryan DSM, Marquardt JPCM, Wrobel MM, Bridge CP, et al. Multilevel body composition analysis on chest computed tomography predicts hospital length of stay and complications after lobectomy for lung cancer: a multicenter study. *Ann Surg* 2020.
  19. Blanc-Durand P, Campedel L, Mule S, Jegou S, Luciani A, Pigneur F, et al. Prognostic value of anthropometric measures extracted from whole-body CT using deep learning in patients with non-small-cell lung cancer. *Eur Radiol* 2020.
  20. Bridge CP, Rosenthal M, Wright B, Kotecha G, Fintelmann F & Troschel F Fully-automated analysis of body composition from CT in cancer patients using convolutional neural networks. *ArXiv180803844 Cs* 2018;11041:204–213. [https://doi.org/10.1007/978-3-030-01201-4\\_22](https://doi.org/10.1007/978-3-030-01201-4_22)
  21. Hemke R, Buckless CG, Tsao A, Wang B, Torriani M. Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment. *Skeletal Radiol* 2020;**49**(3):387–395. <https://doi.org/10.1007/s00256-019-03289-8>
  22. Magudia K, Bridge CP, Bay CP, Babic A, Fintelmann FJ, Troschel FM, et al. Population-scale CT-based body composition analysis of a large outpatient population using deep learning to derive age-, sex-, and race-specific reference curves. *Radiology* 2021;**298**(2):319–329. <https://doi.org/10.1148/radiol.2020201640>
  23. Bayar MA, Antoun S, Lanoy E. Statistical approaches for evaluating body composition markers in clinical cancer research. *Expert Rev Anticancer Ther* 2017;**17**:311–318.
  24. Grønberg BH, Sjøblom B, Wentzel-Larsen T, Baracos VE, Hjerbstad MJ, Aass NK, et al. A comparison of CT based measures of skeletal muscle mass and density from the Th4 and L3 levels in patients with advanced non-small-cell lung cancer. *Eur J Clin Nutr* 2018;**73**:1–8.
  25. Lage DE, Nipp RD, D'Arpino SM, Moran SM, Johnson PC, Wong RL, et al. Predictors of posthospital transitions of care in patients with advanced cancer. *J Clin Oncol* 2017;**36**:76–82.
  26. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics* 2009;**50**(6):613–621. <https://doi.org/10.1176/appi.psy.50.6.613>
  27. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. *J Palliat Care* 2019.
  28. Chang VT, Hwang SS, Feuerman M. Validation of the Edmonton Symptom Assessment Scale. *Cancer* 2000;**88**:2164–2171.
  29. van Seventer E, Marquardt JP, Troschel AS, Best TD, Horick N, Azoba C, et al. Associations of skeletal muscle with symptom burden and clinical outcomes in hospitalized patients with advanced cancer. *J Natl Compr Canc Netw* 2021;**1**:1–9.
  30. Fuchs G, Chretien YR, Mario J, Do S, Eikermann M, Liu B, et al. Quantifying the effect of slice thickness, intravenous contrast and tube current on muscle segmentation: implications for body composition analysis. *Eur Radiol* 2018;**28**:2455–2463.
  31. Janssen I, Heymsfield SB, Wang Z, Ross R. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. *J Appl Physiol* 2000;**89**:81–88.
  32. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2020 <https://www.R-project.org/>
  33. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;**30**:1105–1117.
  34. Dabiri S, Popuri K, Cespedes Feliciano EM, Caan BJ, Baracos VE, Beg MF. Muscle segmentation in axial computed tomography (CT) images at the lumbar (L3) and thoracic (T4) levels for body composition analysis. *Comput Med Imaging Graph* 2019;**75**:47–55.
  35. Yao L, Petrosyan A, Fuangfa P, Lenchik L, Boutin RD. Diagnosing sarcopenia at the point of imaging care: analysis of clinical, functional, and opportunistic CT metrics. *Skeletal Radiol* 2021;**50**(3):543–550. <https://doi.org/10.1007/s00256-020-03576-9>
  36. Harel O, Zhou X-H. Multiple imputation: review of theory, implementation and software. *Stat Med* 2007;**26**:3057–3077.
  37. Abate N, Garg A, Coleman R, Grundy SM, Peshock RM. Prediction of total subcutaneous abdominal, intraperitoneal, and retroperitoneal adipose tissue masses in men by a single axial magnetic resonance imaging slice. *Am J Clin Nutr* 1997;**65**:403–408.
  38. Irlbeck T, Massaro JM, Bamberg F, O'Donnell CJ, Hoffmann U, Fox CS. Association between single-slice measurements of visceral and abdominal subcutaneous adipose tissue with volumetric measurements: the Framingham Heart Study. *Int J Obes (Lond)* 2010;**34**:781–787.
  39. Troschel F, Jin Q, Eichhorn F, Best T, Leppelmann K, Troschel A, et al. Sarcopenia on preoperative chest computed tomography predicts cancer-specific and all-cause mortality following pneumonectomy for lung cancer: a multicenter analysis. *Cancer Med* in press.
  40. von Haehling S, Morley JE, Coats AJS, Anker SD. Ethical guidelines for publishing in the journal of cachexia, sarcopenia and muscle: update 2019. *J Cachexia Sarcopenia Muscle* 2019;**10**:1143–1145.

## Appendix A

Proof 1:

We state that in calculating the average percentile, not including missing values is mathematically equivalent to silently imputing them as the mean of present variables.

Let the average of values  $x$  and  $y$  be  $a = \frac{x+y}{2}$ . Let  $z_1, z_2, \dots, z_n$  be missing values imputed as the mean of  $x$  and  $y$ , so that  $z = \frac{x+y}{2} = a$ . Let the average calculated after imputation be  $a_i = \frac{x+y+z_1+\dots+z_n}{n+2}$ . Solving by insertion step by step:

$$\begin{aligned} a_i &= \frac{x+y+z_1+\dots+z_n}{n+2} \\ &= \frac{x+y}{n+2} + \frac{z_1+\dots+z_n}{n+2} \\ &= \frac{x+y}{n+2} + \frac{n\left(\frac{x+y}{2}\right)}{n+2} \\ &= \frac{2(x+y)}{2(n+2)} + \frac{n(x+y)}{2(n+2)} \\ &= \frac{(n+2)(x+y)}{2(n+2)} \\ &= \frac{x+y}{2} \\ &= a \end{aligned}$$

qed.

**Table A1** Characteristics of patients screened for imaging, those with skeletal muscle measurements, and those excluded because of lacking muscle measurements

Characteristic	Overall (n = 1121)	Muscle data available (n = 846)	No muscle data available (n = 275)	P value*
Age—mean (SD)	64.07 (12.80)	63.46 (12.90)	65.95 (12.33)	0.005
Sex (%)				0.317
Male	576 (51.4)	427 (50.5)	149 (54.2)	
Female	545 (48.6)	419 (49.5)	126 (45.8)	
Race/ethnicity (%)				0.472
White	1038 (92.6)	779 (92.1)	259 (94.2)	
African American	39 (3.5)	30 (3.5)	9 (3.3)	
Asian	22 (2.0)	2 (0.7)	20 (2.4)	
Hispanic	19 (1.7)	4 (1.5)	15 (1.8)	
Other	3 (0.3)	2 (0.2)	1 (0.4)	
Marital status (%)				0.349
Married	737 (65.7)	557 (65.8)	180 (65.5)	
Single	170 (15.2)	135 (16.0)	35 (12.7)	
Divorced	125 (11.2)	92 (10.9)	33 (12.0)	
Widowed	89 (7.9)	62 (7.3)	27 (9.8)	
Education level (%)				0.954
High school and below	350 (31.2)	266 (31.4)	84 (30.5)	
Beyond high school	672 (59.9)	505 (59.7)	167 (60.7)	
Declined to provide	99 (8.8)	75 (8.9)	24 (8.7)	
Health insurance (%)				0.006
Government-sponsored	587 (52.4)	421 (49.8)	166 (60.4)	
Private	530 (47.3)	421 (49.8)	109 (39.6)	
None	4 (0.4)	4 (0.5)	0 (0.0)	
Cancer type (%)				<0.001
Gastrointestinal	331 (29.5)	278 (32.9)	53 (19.3)	
Lung	188 (16.8)	160 (18.9)	28 (10.2)	
Genitourinary	112 (10.0)	76 (9.0)	36 (13.1)	
Melanoma	90 (8.0)	68 (8.0)	22 (8.0)	
Breast	75 (6.7)	56 (6.6)	19 (6.9)	
Gynaecologic	52 (4.6)	42 (5.0)	10 (3.6)	
Sarcoma	50 (4.5)	45 (5.3)	5 (1.8)	
Head and neck	54 (4.8)	41 (4.8)	13 (4.7)	
Leukaemia	61 (5.4)	35 (4.1)	26 (9.5)	
Lymphoma	94 (8.4)	33 (3.9)	61 (22.2)	
Cancer of unknown primary	14 (1.2)	12 (1.4)	2 (0.7)	
Months since cancer diagnosis—mean (SD)	16.81 (23.78)	15.20 (22.31)	22.39 (27.67)	<0.001
Charlson comorbidity index—mean (SD)	0.89 (1.29)	0.84 (1.25)	1.02 (1.40)	0.051
Body mass index—mean (SD)	25.94 (6.02)	25.87 (5.96)	27.19 (7.07)	0.160

CT, computed tomography; SD, standard deviation.

\*P value for test of intergroup differences. We used t-tests for continuous variables, Fisher's exact test for race/ethnicity, and  $\chi^2$  tests for all other categorical variables.



**Table A2** Comparison of skeletal muscle metrics (rows) and vertebral levels (columns) across concordance statistic, number of included patients, coefficient of determination, and *P* value

Outcome (metric)/vertebral level(s)	PBA			L3			T10			T8			T5			PBA			
	<i>n</i>	<i>C</i>	<i>R</i> <sup>2</sup>	<i>n</i>	<i>C</i>	<i>R</i> <sup>2</sup>	<i>n</i>	<i>C</i>	<i>R</i> <sup>2</sup>	<i>n</i>	<i>C</i>	<i>R</i> <sup>2</sup>	<i>n</i>	<i>C</i>	<i>R</i> <sup>2</sup>	<i>n</i>	<i>C</i>	<i>R</i> <sup>2</sup>	
<b>Performance characteristic</b>																			
<b>SMI-based models</b>																			
OS (SMI)	843	0.62	0.087	2.8e-09	675	0.59	0.072	5.6e-06	782	0.60	0.076	1.0e-05	672	0.59	0.070	4.0e-06	576	0.63	0.120
RD (SMI)	843	0.60	0.072	3.2e-07	675	0.60	0.051	2.4e-03	782	0.60	0.064	1.7e-04	672	0.59	0.066	1.2e-05	576	0.61	0.095
RD90 (SMI)	843	0.61	0.028	1.2e-04	675	0.58	0.014	3.1e-01	782	0.60	0.022	1.3e-02	672	0.62	0.032	1.5e-04	576	0.63	0.030
LOS (SMI)	843	0.56	0.039	4.6e-02	675	0.55	0.026	6.9e-02	782	0.55	0.039	2.3e-01	672	0.55	0.045	2.0e-01	576	0.57	0.065
PHQ4D (SMI)	698	0.58	0.039	2.7e-02	563	0.59	0.044	1.8e-01	644	0.58	0.039	4.3e-02	559	0.59	0.041	4.3e-02	477	0.58	0.036
PHQ4A (SMI)	701	0.58	0.042	3.4e-01	565	0.58	0.038	4.4e-01	647	0.58	0.046	8.2e-02	562	0.59	0.049	5.8e-01	479	0.59	0.043
ESAsT (SMI)	827	0.58	0.054	1.1e-01	662	0.59	0.065	1.5e-01	768	0.57	0.047	1.4e-01	664	0.58	0.053	1.5e-01	569	0.58	0.059
ESAsP (SMI)	833	0.57	0.045	1.2e-01	667	0.58	0.054	1.8e-01	774	0.57	0.041	1.7e-01	669	0.57	0.049	1.7e-01	573	0.57	0.055
<b>SMRA-based models</b>																			
OS (SMRA)	843	0.61	0.094	1.6e-10	675	0.60	0.081	1.7e-07	782	0.60	0.093	1.7e-08	672	0.60	0.071	7.7e-06	576	0.61	0.084
RD (SMRA)	843	0.61	0.077	5.3e-08	675	0.61	0.069	2.4e-06	782	0.60	0.088	1.1e-08	672	0.60	0.059	1.1e-04	576	0.60	0.063
RD90 (SMRA)	843	0.62	0.030	4.9e-05	675	0.61	0.025	1.5e-03	782	0.63	0.036	1.6e-05	672	0.62	0.027	2.1e-03	576	0.60	0.020
LOS (SMRA)	843	0.56	0.049	3.3e-04	675	0.56	0.035	3.4e-03	782	0.55	0.046	6.5e-03	672	0.55	0.059	2.6e-03	576	0.56	0.064
PHQ4D (SMRA)	698	0.59	0.050	3.1e-04	563	0.60	0.053	6.2e-03	644	0.59	0.051	5.9e-04	559	0.58	0.038	1.5e-01	477	0.58	0.040
PHQ4A (SMRA)	701	0.59	0.053	2.3e-03	565	0.58	0.049	1.0e-02	647	0.59	0.055	2.7e-03	562	0.59	0.052	1.7e-01	479	0.59	0.044
ESAsT (SMRA)	827	0.58	0.063	1.6e-03	662	0.59	0.080	2.2e-03	768	0.58	0.055	1.4e-02	664	0.58	0.058	1.6e-02	569	0.58	0.064
ESAsP (SMRA)	833	0.57	0.047	4.7e-02	667	0.58	0.061	1.8e-02	774	0.57	0.043	2.2e-01	669	0.57	0.048	2.2e-01	573	0.57	0.052
<b>SMG-based models</b>																			
OS (SMG)	843	0.62	0.103	1.7e-12	675	0.60	0.090	7.5e-09	782	0.61	0.103	8.6e-10	672	0.60	0.083	1.1e-07	576	0.63	0.120
RD (SMG)	843	0.61	0.082	3.6e-09	675	0.61	0.072	1.0e-06	782	0.60	0.094	2.0e-09	672	0.59	0.074	8.0e-07	576	0.61	0.093
RD90 (SMG)	843	0.62	0.032	1.4e-05	675	0.60	0.023	3.7e-03	782	0.63	0.039	4.1e-06	672	0.64	0.039	1.0e-05	576	0.63	0.031
LOS (SMG)	843	0.56	0.050	2.0e-04	675	0.56	0.035	4.3e-03	782	0.56	0.044	1.5e-02	672	0.55	0.056	8.9e-03	576	0.57	0.070
PHQ4D (SMG)	698	0.59	0.048	7.3e-04	563	0.60	0.050	1.7e-02	644	0.59	0.050	8.8e-04	559	0.59	0.042	4.1e-02	477	0.58	0.038
PHQ4A (SMG)	701	0.59	0.048	1.9e-02	565	0.58	0.045	4.4e-02	647	0.59	0.053	5.8e-03	562	0.58	0.051	3.0e-01	479	0.59	0.043
ESAsT (SMG)	827	0.58	0.062	2.1e-03	662	0.59	0.079	2.7e-03	768	0.58	0.056	8.4e-03	664	0.58	0.061	6.8e-03	569	0.58	0.065
ESAsP (SMG)	833	0.57	0.048	2.1e-02	667	0.58	0.063	9.2e-03	774	0.57	0.045	7.4e-02	669	0.57	0.052	4.0e-02	573	0.58	0.057

*C*, concordance statistic; ESAsP, Edmonton Symptom Assessment System physical; ESAsT, Edmonton Symptom Assessment System total; L3, third lumbar; LOS, hospital length of stay; *n*, number of included patients; OS, overall survival; *P*, *P* value; PBA, percentile-based averaging; PHQ4A, patient health Questionnaire-4 anxiety; PHQ4D, patient health Questionnaire-4 depression; *R*<sup>2</sup>, coefficient of determination; RD, time to unplanned hospital readmissions or death; RD90, readmission or death within 90 days; SM, skeletal muscle; SMG, skeletal muscle gauge; SMI, skeletal muscle index; SMRA, skeletal muscle radioattenuation; T5/T8/T10, fifth/eighth/tenth thoracic.