Nomograms for Automated Body Composition Analysis:
A Crucial Step for Routine Clinical Implementation

Ronald M. Summers, MD, PhD

Dr Summers is a tenured senior investigator and staff radiologist at the NIH Clinical Center. His research interests include large radiology image databases, machine learning, and abdominal imaging. He is a fellow of the Society of Abdominal Radiologists and of the American Institute for Medical and Biological Engineering, a member of several journal editorial boards, and a recipient of the Presidential Early Career Award for Scientists and Engineers.

Body composition (BC) analysis is the measurement of fat, muscle, and bone for assessment of fitness and health. Analysis of BC on radiologic images has a long tradition. According to PubMed, there are more than 1400 articles on BC and CT, with the first appearing shortly after the commercialization of CT (1). There are good reasons for this interest. BC is predictive of clinical outcome in a number of important scenarios. Examples include risk prediction for patients undergoing surgery or chemotherapy, mortality and cardiovascular morbidity prediction, and assessment of metabolic syndrome and adverse drug effects.

Body mass index is one method of determining BC but is too simplistic. Hence the need for imaging-based methods. Until recently, BC was measured manually using simple techniques such as thresholding and manual tracing on a single abdominal CT or MRI scan, typically between the L1 and L4 levels. Because manual assessment of BC is time consuming and expensive, its use has been limited to research studies and clinical trials. More recently, partially and fully automated software for measuring BC has been reported. The study by Magudia et al in this issue of *Radiology* describes one such automated assessment (2).

Magudia et al have developed a deep learning BC analysis pipeline for routine abdominal CT. Their software measures muscle and visceral and subcutaneous fat at the L3 level. The software automatically finds the L3 level and then performs single-slice fat and muscle analyses. The authors developed their software on 595 scans from patients with pancreatic adenocarcinoma from a multi-institution study and 534 scans from patients with lymphoma from a different institution. They validated the results using an additional 100 manually segmented images. Finally, they applied the software to the CT scans of 12 128 outpatients from three hospitals to generate age-, race-, and sex-normalized reference curves or nomograms. As expected, normal BC varied significantly with age, race, and sex. The authors also found that skeletal muscle area z scores (a normalized measure) were significant predictors of 2-year survival in combined models that included body mass index.

What are the implications of this and similar studies? For one, routine quantification of abdominal CT for BC could be done on every scan fully automatically and at little or no additional cost. The data so acquired could be combined with other clinical information such as patient history, laboratory values, and drug treatments to make predictions that are clinically relevant for the particular patient. The nomograms presented in the article by Magudia et al are important because the BC can best be properly assessed and combined with other clinical data when they are corrected for age, sex, and race.

BC metrics are some of the easiest to automate in an accurate way and tend to be more robust than other current machine learning–based techniques. This is because fat has a well-defined Hounsfield unit range and muscle anatomic variation at the L3 level tends to be relatively limited and easily modeled using small training datasets. BC also has a strong track record of clinical relevance across numerous published studies. The combination of ease of implementation, robustness, and proven clinical utility suggests that automated BC could become routinely available clinically sooner rather than later.

Some of the limitations of the current work include a lack of analysis according to whether intravenous contrast material was given, limitations in the number of races investigated (only White non-Hispanic and Black patients were included), analyses at only a single vertebral level, and a reference population of outpatients who cannot be regarded as healthy and who could have atypical BCs. Recent work indicates that muscle, bone, and fat measurements vary between unenhanced and intravenous contrast material–enhanced CT of the same patient (3). Such measurements can be normalized by means of a linear correction if the data are first stratified accordingly by whether contrast material is given. While Magudia et al made their software publicly available, they did not publish their trained model, CT datasets, and manual segmentations, thereby limiting reproducibility.

The evolution of BC is another topic of interest. Interval changes are important in treatments for obesity, assessment of renal and bone marrow transplant recipients, and population-based assessments of nutritional status. There has been limited work in the use of automated BC software for longitudinal analyses.

What is one to do with the blizzard of data presented in Magudia et al's supplementary material? In all likelihood, these nomograms would be incorporated into the software and used to compute the z scores without the need for the radiologist to consult the tables. These z scores would incorporate the age-, sex-, and race-specific distributions found by the authors. One would then know where an individual patient stands in terms of BC. A 2-year survival prediction could also be output. A summary of the findings could be inserted directly into the radiology report, requiring the radiologist only to double check whether the results seem reasonable for a given patient.

The radiologist's intuition about the reasonableness of the results would likely develop quickly if the software were applied on every case. Radiologists who interpret images from dual-energy x-ray absorptiometry will already be familiar with normalized scores. Radiologists who perform US on children will already be familiar with the use of nomograms for assessing organ size as a function of age and sex. These are not new concepts, just applied for a new indication.

How do the authors' findings compare with others in the literature? In a recent study from my collaboration with Dr Perry Pickhardt at the University of Wisconsin, we found that BC assessment of both muscle and visceral-to-subcutaneous fat ratio were good predictors of 2-year survival in patients undergoing colorectal cancer screening (4). Other work from our collaboration showed that automated BC was predictive of metabolic syndrome and future osteoporotic fractures (5,6). It is clear that automated systems such as the one presented by Magudia et al will prove useful for prediction of important clinical outcomes.

How do the authors' segmentation results compare with other automated BC analyses in the literature? Weston and colleagues (7) found Dice scores (a measure of segmentation accuracy) of 0.97 and 0.98 for visceral and subcutaneous fat, respectively, and 0.96 for muscle assessments. Burns et al (8) showed a Dice score of 0.94 for muscle segmentation. These results are comparable to the Dice score range of 0.95–0.98 reported by Magudia et al and indicate that such analyses are reproducible across different deep learning methods. However, more data are needed to show generalizability in multi-institutional datasets.

What is the likely future for automated BC assessment? Analyses of tissues other than fat and muscle at abdominal imaging are being automated and validated. Although these tissues are not traditionally considered part of BC, these analyses fall roughly into the nascent category of opportunistic screening, the identification of abnormalities other than those addressing the primary indication for the scan. Examples include bone mineral densitometry for osteoporosis, noncoronary calcific atherosclerosis scoring for cardiovascular and oncology risk prediction, and liver steatosis for nonalcoholic fatty liver assessment. One can imagine other automated assessments being routinely performed, such as major organ volumetrics (liver, spleen, kidneys, pancreas) and management of incidentalomas. In short, BC analyses such as the work by Magudia et al are likely to lead the way to more routine clinical use of automated analyses in abdominal imaging.

References

1. Huang HK, Wu SC. The evaluation of mass densities of human body in vivo from CT scans. Comput Biol Med 1976;6(4):337–343.
2. Magudia K, Bridge CP, Bay CP, 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;298:319–329.
3. Perez AA, Pickhardt PJ, Elton DC, Sandfort V, Summers RM. Fully automated CT imaging biomarkers of bone, muscle, and fat: correcting the effect of intravenous contrast. Abdom Radiol (NY) doi: 10.1007/s00261-020-02755-5. Published online September 18, 2020.
4. Pickhardt PJ, Graffy PM, Zea R, 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(4):e192–e200.
5. Pickhardt PJ, Graffy PM, Zea R, et al. Opportunistic Screening for Metabolic Syndrome in Asymptomatic Adults Utilizing Fully Automated Abdominal CT-based Biomarkers. AJR Am J Roentgenol doi: 10.2214/AJR.20.23049. Published online June 29, 2020.
6. Pickhardt PJ, Graffy PM, Zea R, et al. Automated Abdominal CT Imaging Biomarkers for Opportunistic Prediction of Future Major Osteoporotic Fractures in Asymptomatic Adults. Radiology 2020;297(1):64–72.
7. Weston AD, Korfiatis P, Kline TL, et al. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. Radiology 2019;290(3):669–679.
8. Burns JE, Yao J, Chalhoub D, Chen JJ, Summers RM. A Machine Learning Algorithm to Estimate Sarcopenia on Abdominal CT. Acad Radiol 2020;27(3):311–320.