

## **Derivation of anthropometric-based equations to predict lean body mass composition of cancer patients**

Autumn B. Carey<sup>1</sup>, Ashley S. Felix<sup>1</sup>, Jared D. Huling<sup>2</sup>, James B. Odei<sup>3</sup>, Christopher C. Coss<sup>4</sup>, Xia Ning<sup>5, 6</sup>, and Macarius M. Donneyong<sup>7, 8</sup>

<sup>1</sup>*Division of Epidemiology, The Ohio State University College of Public Health, Columbus, OH 43210*

<sup>2</sup>*Division of Biostatistics, University of Minnesota School of Public Health, Minneapolis, MN 55455*

<sup>3</sup>*Division of Biostatistics, The Ohio State University College of Public Health, Columbus, OH 43210*

<sup>4</sup>*Department of Pharmaceutics and Pharmaceutical Chemistry, The Ohio State University College of Pharmacy, Columbus, OH 43210*

<sup>5</sup>*Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, OH 43210*

<sup>6</sup>*Department of Computer Science and Engineering, The Ohio State University College of Engineering, Columbus, OH 43210*

<sup>7</sup>*Department of Pharmacy Practice and Science, The Ohio State University College of Pharmacy, Columbus, OH 43210*

<sup>8</sup>*Division of Health Services Management and Policy, The Ohio State University College of Public Health, Columbus, OH 43210*

### **CORRESPONDING AUTHOR:**

Macarius M. Donneyong, PhD, MPH

Riffe Building, 400

500 W. 12<sup>th</sup> Ave.

Columbus, OH 43210

[donneyong.1@osu.edu](mailto:donneyong.1@osu.edu)

**Disclosure:** The authors do not have any disclosures relevant to this study.

1 **ABSTRACT**

2 Background: Lean body mass (LBM) composition of cancer patients is a predictor of  
3 chemotherapy-related adverse events and overall cancer survival. However, clinicians  
4 lack validated algorithms that can be applied to measure the LBM of cancer patients to  
5 facilitate accurate chemotherapy dosing. Our goal was to develop LBM predictive  
6 equations using routinely measured anthropometric measures among cancer patients.  
7

8 Methods: We leveraged the 1999-2006 National Health and Nutrition Examination Survey  
9 (NHANES) data cycles containing information on self-reported cancer diagnosis, LBM  
10 measures based on dual-energy x-ray absorptiometry (DXA) and several anthropometric  
11 and demographic factors. We restricted our analysis to participants who had been  
12 diagnosed with cancer at the time of surveys. The data was randomly split to 75%:25%  
13 to train and test predictive models. Least absolute shrinkage and selection operator  
14 (LASSO) models were used to predict LBM based on anthropometric and demographic  
15 factors, overall and separately among sex and sex-by-race/ethnic subgroups. LBM  
16 measured directly with DXA served as the gold standard for assessing the predictive  
17 abilities (correlations [ $R^2$ ] and the Root Mean Square Error [RMSE]) of the derived LBM-  
18 algorithms. We further compared the correlations between both DXA-based LBM and  
19 predicted LBM and urine creatinine levels, a known biomarker of muscle mass.  
20

21 Results: We identified 1,777 cancer patients with a median age of 71 (interquartile range  
22 [IQR]: 60-80) years. The most parsimonious model comprised of height and weight, which  
23 accurately predicted LBM overall ( $R^2=0.86$ , RMSE =2.26). The predictive abilities of these  
24 models varied across sex-by-race/ethnic groups. The magnitude of correlations between  
25 derived LBM-algorithm and urine creatinine levels were larger compared to those  
26 measured between DXA-based LBM and urine creatinine levels ( $R^2=0.30$  vs.  $R^2=0.17$ )  
27

28 Conclusions: We successfully developed a simple sex-specific and sex-by-race/ethnicity-  
29 specific models to accurately predict the LBM of cancer patients by using only height and  
30 weight. The simplicity and high accuracy of these models make them inexpensive  
31 alternatives to measuring the LBM of cancer patients. Data on the LBM of cancer patients  
32 could help guide optimal chemotherapy dose selection among cancer patients.

## 33 INTRODUCTION

34 In 2020, approximately 1.8 million cancers are expected to be diagnosed in the  
35 United States, with approximately 0.6 million cancer-related deaths<sup>1</sup>. Cancer is the  
36 second leading cause of death with a five-year mortality rate of 34.1%<sup>1,2</sup>. Notwithstanding  
37 this high mortality burden, advances in treatment underlie significant declines in cancer  
38 mortality over the past 20 years, with rates dropping from 209.9 per 100,000 people in  
39 1995 to 158.7 per 100,000 people in 2015<sup>2</sup>. Chemotherapy has been shown to improve  
40 cancer survival, especially for advanced stage cancers<sup>3,4</sup>. However, this benefit could be  
41 offset by the negative impact of chemotherapy-induced adverse events and chemotoxicity  
42 effects, such as hair loss, nausea, vomiting, or death. These adverse effects can lead to  
43 treatment cycle interruption or discontinuation, both of which contribute to poor cancer  
44 outcomes and higher mortality<sup>5</sup>.

45 Chemotoxicity is caused by the excretion or poor metabolism of chemotherapeutic  
46 drugs, resulting in higher than expected circulating levels of the drug<sup>6,7</sup>. Known risk  
47 factors for poor chemotherapeutic drug metabolism are determined by the  
48 pharmacokinetic properties of the chemotherapeutic drug, but often include low lean  
49 muscle mass<sup>6</sup>. Prado et al (2007) found that a cut point of 20mg 5-fluorouracil per  
50 kilogram of total lean body mass (LBM) is a threshold for developing toxicity (p-  
51 value<0.01)<sup>6</sup>. LBM is comprised of metabolic tissue (liver and kidneys), skeletal muscle,  
52 intracellular water, extracellular water, and bone<sup>8</sup>. Thus, research continues to emphasize  
53 the value of measuring body composition to normalize drug dosing, optimize outcomes,  
54 and mitigate chemotoxicity<sup>6,9-12</sup>.

55 In clinical practice, chemotherapy dosing is based on body weight, body mass  
56 index (BMI) or body surface area (BSA). This practice may pose challenges to measuring  
57 the optimal dose of chemotherapy for individual patients. BMI is limited as it does not  
58 capture a valid measure of body composition phenotypes (i.e. lean muscle mass, lean  
59 muscle mass percentage, fat mass, and fat mass percentage) and does not readily  
60 explain inter-patient variability in drug clearance inter-patient variability despite its use in  
61 many epidemiological studies assessing chemotoxicity risk and cancer outcomes<sup>13,14</sup>.  
62 Chemotherapy dosing based on BSA is a common clinical practice in oncologic settings.

63 However, variation in the pharmacokinetics of most chemotherapies has failed to be  
64 standardized by BSA-based dosing<sup>15-17</sup>.

65 Dual-energy x-ray absorptiometry (DXA) and computed tomography (CT) are  
66 imaging tools used to accurately calculate body composition among cancer patients.  
67 However, CT scans are expensive and emit a high dose of ionizing radiation<sup>18</sup>. DXA is  
68 relatively inexpensive and emits low amounts of radiation, but whole body DXA  
69 examination equipment is not typically accessible in an oncologic clinic setting<sup>18, 19</sup>. Due  
70 to these limitations, anthropometric measures in conjunction with age, sex, weight, and  
71 height are needed to precisely estimate body composition for chemotoxicity risk  
72 stratification<sup>13, 20</sup>.

73 Several studies have shown that LBM can be estimated accurately by using  
74 equations that comprise of simple anthropometric measures<sup>21-24</sup>. However, none of these  
75 studies published data the application of anthropometric-based equations to measure  
76 LBM specifically among cancer patients. Therefore, our primary objective was to develop  
77 anthropometric-based models to accurately predict LBM among cancer patients.

## 78 **METHODS**

### 79 ***Study Population***

80 Participants were selected from the National Health and Nutrition Survey  
81 (NHANES), a series of cross-sectional surveys designed to assess the health and  
82 nutritional status of adults and children in the United States conducted by the Centers for  
83 Disease Control and Prevention<sup>25</sup>. Because NHANES is a publically available dataset,  
84 The Ohio State University IRB did not constitute the project to require human subjects  
85 review. NHANES participants were selected using complex, stratified, multi-staged  
86 probability sampling to be representative of the US civilian non-institutionalized  
87 population<sup>26-28</sup>. All survey cycles between 1999 and 2014 were utilized. **Figure 1** depicts  
88 sample selection for the development of the final training, testing and validation of models.

### 89 **Inclusion/Exclusion criteria**

90 We included participants with a self-reported cancer diagnosis aged  $\geq 18$  years old.  
91 Pregnant women were excluded<sup>29</sup>.

### 92 ***Outcomes***

#### 93 **Dual-Energy X-Ray Absorptiometry Measurements**

94 Whole body DXA scans were used to measure LBM, fat mass, and fat mass  
95 percent of NHANES participants. A Hologic QDR 4500A fan-beam densitometer within  
96 the NHANES Mobile Examination Centers was used to measure body composition<sup>30-32</sup>.  
97 Details of the body composition measurement and exclusion of participants have been  
98 previously published<sup>31</sup>. Outliers at the individual level were coded as missing by NHANES  
99 staff prior to completing statistical analysis<sup>32</sup>.

## 100 **Predictors**

### 101 **Anthropometric Measurements**

102 We focused on the following anthropometric measures that have been assessed  
103 as predictors of LBM in the published data <sup>21-24</sup>: height (cm), weight (kg), arm  
104 circumference (cm), abdominal circumference (cm), calf circumference (cm), thigh  
105 circumference (cm), tricep skinfold (cm), and subscapular skinfold (cm). These  
106 anthropometric measures were collected in the Mobile Examination Center by two trained  
107 health technicians, an examiner and recorder, per NHANES standardized procedures and  
108 equipment<sup>33</sup>. For the modeling procedures, all anthropometric variables were modeled as  
109 continuous variables, and tricep and subscapular skinfold measures were converted from  
110 units of millimeters to centimeters.

### 111 **Other measures**

112 Demographic factors examined in this study are age, time between cancer  
113 diagnosis and NHANES screening age (years), race (Black Non-Hispanic, Mexican  
114 American and Hispanic, White Non-Hispanic and other (which includes multiracial), and  
115 sex (male and female). Age was top coded at 80 years old in NHANES cycles 1999-2006  
116 and 85 years old in NHANES cycles 2007-2014 to maintain confidentiality. Thus, for this  
117 analysis, all individuals over the age of 80 were top coded at 80 to be uniform across all  
118 survey cycles. Additionally, age of cancer diagnosis or most recent cancer diagnosis, if  
119 multiple diagnoses were reported, were examined to calculate the time between cancer  
120 diagnoses and LBM measurement. In addition to these demographic factors, we selected  
121 urine creatinine levels as a potential biomarker for muscle <sup>34, 35</sup>. Urine creatinine levels  
122 were measured via 24-hour urine samples collected within 2 weeks of DXA  
123 measurements and analyzed by NHANES staff <sup>36</sup>.

### 124 **Statistical Analysis**

125 The complex sampling design of NHANES was accounted for by proper use of the  
126 sample and cycle weights in all analysis<sup>37</sup>. We implemented the analysis described  
127 below with SAS 9.4 and R 1.2.1335.

## 128 **Descriptive analysis**

129 Proportions and means were used to describe categorical and continuous  
130 variables, respectively, overall, by sex, and by NHANES data cycles (with and without  
131 whole-body DXA measurements).

## 132 **Missing Data**

133 Greater than 10% of participants included in our analytic sample were missing data  
134 for BMI (kg/m<sup>2</sup>), BSA (m<sup>2</sup>), all eight anthropometric measures, creatinine (mg/dL), and  
135 DXA measures. We applied Multivariate Imputation by Chained Equations (MICE)  
136 models<sup>38</sup> to impute all missing values of predictors. MICE requires missing data models  
137 to perform the imputation, for which flexible random forest models were used due to their  
138 favorable performance<sup>39</sup>. To obtain valid estimates that take into account the uncertainty  
139 in imputation, we constructed twenty-five imputed datasets were constructed. NHANES  
140 had already imputed missing DXA values prior to making data available to researchers  
141 by using multiple-imputation methodology by the National Center for Health Statistics  
142 (NCHS)<sup>32</sup>. This missing data imputation was based on a sequential regression  
143 multivariate imputation methodology to reflect smaller standard errors, narrower  
144 confidence intervals, and falsely failing to reject null hypotheses that are often reflected  
145 when utilizing single imputation methodsDXA technical doc<sup>32</sup>.

## 146 **Algorithm development**

147 We implemented five different models based on a combination of anthropometric  
148 and demographic factors among the overall sample: 1) Model 1: height (cm) and weight  
149 (kg) only; 2) Model 2: Model 1 plus age, time between cancer diagnosis and LBM  
150 measurement date, sex, race/ethnicity; 3) Model 3: Model 2 plus interaction terms  
151 between anthropometric predictors (height (cm) and weight (kg)) and sex and age; 4)  
152 Model 4: Model 3 plus additional interaction terms between anthropometric variables and  
153 race/ethnicity; 5) Model 5: Model 4 plus all pairwise interactions between anthropometric  
154 variables, age, time between cancer diagnosis and LBM measurement, sex, and  
155 race/ethnicity.

156 Sex specific models were generated as follows: 1) Model 1: height (cm) and weight  
157 (kg) only; 2) Model 2: Model 1 plus age, time between cancer diagnosis and LBM  
158 measurement date, and race/ethnicity; 3) Model 3: Model 2 plus interaction terms  
159 between anthropometric predictors (height (cm) and weight (kg)) and age; 4) Model 4:  
160 Model 3 plus additional interaction terms between anthropometric variables and  
161 race/ethnicity; 5) Model 5: Model 4 plus all pairwise interactions between anthropometric  
162 variables, age, time between cancer diagnosis and LBM measurement, sex, and  
163 race/ethnicity.

164 To generate sex-by-race/ethnicity-specific predictive models, we constructed four  
165 different models among sex-by-race/ethnicity subgroups: 1) Model 1: height (cm) and  
166 weight (kg) only; 2) Model 2: Model 1 plus age, time between cancer diagnosis and LBM  
167 measurement date; 3) Model 3: Model 2 plus interaction terms between anthropometric  
168 predictors (height (cm) and weight (kg)) and age; 4) Model 5: Model 3 plus all pairwise  
169 interactions between anthropometric variables, age, time between cancer diagnosis and  
170 LBM measurement. Beyond height and weight, we also included other anthropometric  
171 measures in the models specified above in sensitivity analysis to test whether the  
172 inclusion of these other anthropometric variables would yield higher prediction accuracy.

### 173 **Training and testing of algorithms**

174 Data were split randomly into 75%: 25% for training and testing, respectively. For  
175 each imputed dataset, a series of linear regression models with least absolute shrinkage  
176 and selection operator (LASSO) penalty for variable selection were fitted, with a bootstrap  
177 procedure to obtain confidence intervals and p-values<sup>40</sup>. The LASSO was used to develop  
178 accurate but parsimonious predictive equations<sup>40</sup>. Results were aggregated across the  
179 twenty-five imputed datasets to obtain final coefficient estimates. The combination rule  
180 defined by Schomaker et. Al. (2017) was used to combine the bootstrap results across  
181 all 25 imputed datasets and construct (95%) confidence intervals (CIs) that take into  
182 account both the variability of variable selection and imputation<sup>41</sup>. To measure model fit,  
183 root mean square errors (RMSE) and correlation coefficients ( $R^2$ ) were calculated on the  
184 testing data for the overall sample as well as by sex and race/ethnicity subgroups. The  
185 entire training and testing procedure was repeated 10 times and the test set RMSE and  
186  $R^2$  was averaged across all 10 replications for each model.

## 187 **Correlations between predicted biomarkers of muscle mass and LBM**

188 We tested and compared the strengths of correlations ( $R^2$ ) between urine  
189 creatinine levels and predicted LBM vs that between urine creatinine levels and DXA-  
190 based LBM.

## 191 **RESULTS**

### 192 **Study Sample Population Characteristics**

193 The distribution of participant characteristics in the overall study population and by  
194 sex are presented in **Table 1**. We identified 1777 participants with a self-reported cancer  
195 diagnosis with a median age of 71 (IQR:60 – 80). There were 933 (52.5%) females and  
196 844 (47.5%) males. About 23.2% of the population were of a racial/ethnic minority  
197 background: Black (n=218, 12.3%), Mexican American and Hispanic (n=162, 9.1%), and  
198 Other (n=33, 1.9%). Men (55.9kg, SD=9.5) had higher DXA-based LBM than women  
199 (41.1kg, SD=7.4). Men were heavier (83.8kg vs. 73.1kg) and taller than women (174.3cm  
200 vs. 160.7cm).

### 201 **Algorithm Performance**

202 **Table 2** summarizes the performance of the primary model series overall, by sex,  
203 and sex by race stratified models based on  $R^2$  and RMSE values. In the overall sample,  
204 Model 1 was considered the most parsimonious model with high predictive accuracy  
205 ( $R^2=0.86$  and RMSE=4.26). Also, Model 1 had the best prediction accuracy with the  
206 fewest number of predictors (2) among females, ( $R^2=0.93$  and RMSE=2.46) and males,  
207 ( $R^2=0.94$  and RMSE=3.39) and among sex-by-race/ethnicity subgroups.

### 208 **Algorithms**

209 **Table 3** shows the regression coefficients, of the most parsimonious model to  
210 predict DXA-based LBM (kg), overall, by sex and sex-by-race/ethnicity subgroups.  
211 Predictors for the overall model consist of weight (kg) ( $\beta=0.36$ ; p-value<0.001) and height  
212 (cm) ( $\beta=0.52$  p-value<0.05). For females, predictors for the most parsimonious model are  
213 weight (kg) ( $\beta=0.32$ ; p-value<0.001) and height (cm) ( $\beta=0.17$ ; p-value=0.15). For males,  
214 predictors for the most parsimonious model are weight (kg) ( $\beta=0.41$ ; p-value<0.001),  
215 height (cm) ( $\beta=0.32$ ; p-value=0.12). The most parsimonious model for Black, non-  
216 Hispanic females, White, non-Hispanic females, Mexican-American or Hispanic females,  
217 other females, Black, non-Hispanic males, White, non-Hispanic males, Mexican-

218 American or Hispanic males and Other, non-Hispanic males included the same  
219 predictors. Similar results were observed when additional anthropometric variables were  
220 included to build a series of secondary algorithms.

221 The correlations between urine creatinine levels and LBMs (DXA-based and  
222 predicted) are visually represented in **Figure 2**. The magnitude of correlations between  
223 derived LBM-algorithm and urine creatinine levels were large compared to those  
224 measured between DXA-based LBM and urine creatinine levels ( $R^2=0.30$  vs.  $R^2=0.17$ )  
225 The correlations between predicted LBM measures with urine creatinine levels were  
226 comparable to those observed for DXA-based LBM among females ( $R^2=0.25$  vs  $R^2=0.23$ )  
227 and among males ( $R^2=0.17$  vs  $R^2=0.17$ ). These correlations were similar among sex-by-  
228 race/ethnicity subgroups. Therefore, validating our conclusions that the sex by  
229 race/ethnicity stratified Model 1 derived algorithm of LBM (kg) is a sufficient measure for  
230 estimating DXA measured LBM (kg).

## 231 **DISCUSSION**

232 In this large population-based study, we developed and validated predictive  
233 algorithms comprised of simple and routinely measured anthropometric measurements  
234 among cancer patients. Sex, race/ethnicity, and time between cancer diagnosis and DXA  
235 measurement were also added as predictors to improve accuracy among these  
236 subgroups of cancer patients. To the best of our knowledge, there are no previously  
237 published data on the derivation of algorithms for predicting LBM among cancer patients  
238 using routinely measured anthropometric measurements. The developed algorithms will  
239 provide clinicians with a tool to estimate the total LBM composition of cancer patients prior  
240 to initiation of chemotherapy based on routinely measured anthropometric measures  
241 which are non-invasive and inexpensive.

242 In the non-cancer population, several algorithms to predict body compositions  
243 including LBM have been developed and validated<sup>23, 24</sup>. Lee et. al. (2017) recently  
244 developed predictive equations using data from NHANES participants (with or without  
245 cancer diagnosis)<sup>23</sup>. In this study, a statistical model including age, height, weight, waist  
246 circumference, arm circumference, calf circumference, thigh circumference, triceps  
247 skinfold, subscapular skinfold, race, accurately predicted LBM in men ( $R^2=0.94$ ) and  
248 women ( $R^2=0.87$ )<sup>23</sup>. In comparison, our most parsimonious model accurately predicted

249 LBM of cancer patients with the same number of features and to a similar degree of fit as  
250 Lee et al (2017)<sup>23</sup>. Finally, the predictors of the models were limited to the anthropometric  
251 measures collected in NHANES. Published literature supports the use of tricep and  
252 subscapular skinfold in conjunction with circumference measures to accurately measure  
253 body composition <sup>42</sup>.

## 254 **Conclusion**

255 Prediction equations utilizing anthropometric predictors of height and weight to  
256 estimate DXA-based LBM composition of cancer patients were developed, tested and  
257 validated. Models were further improved with sex, race/ethnicity, and time interval  
258 between cancer diagnosis and LBM measurement as predictors. These equations will  
259 simplify the ability of clinicians to estimate the DXA-based LBM composition of cancer  
260 patients prior to initiation of chemotherapy based on routinely measured anthropometric  
261 measures which are non-invasive and low cost to patients. Our next step is to apply these  
262 derived equations to measure LBM thresholds associated with the risk of chemotherapy  
263 adverse events.

## REFERENCES

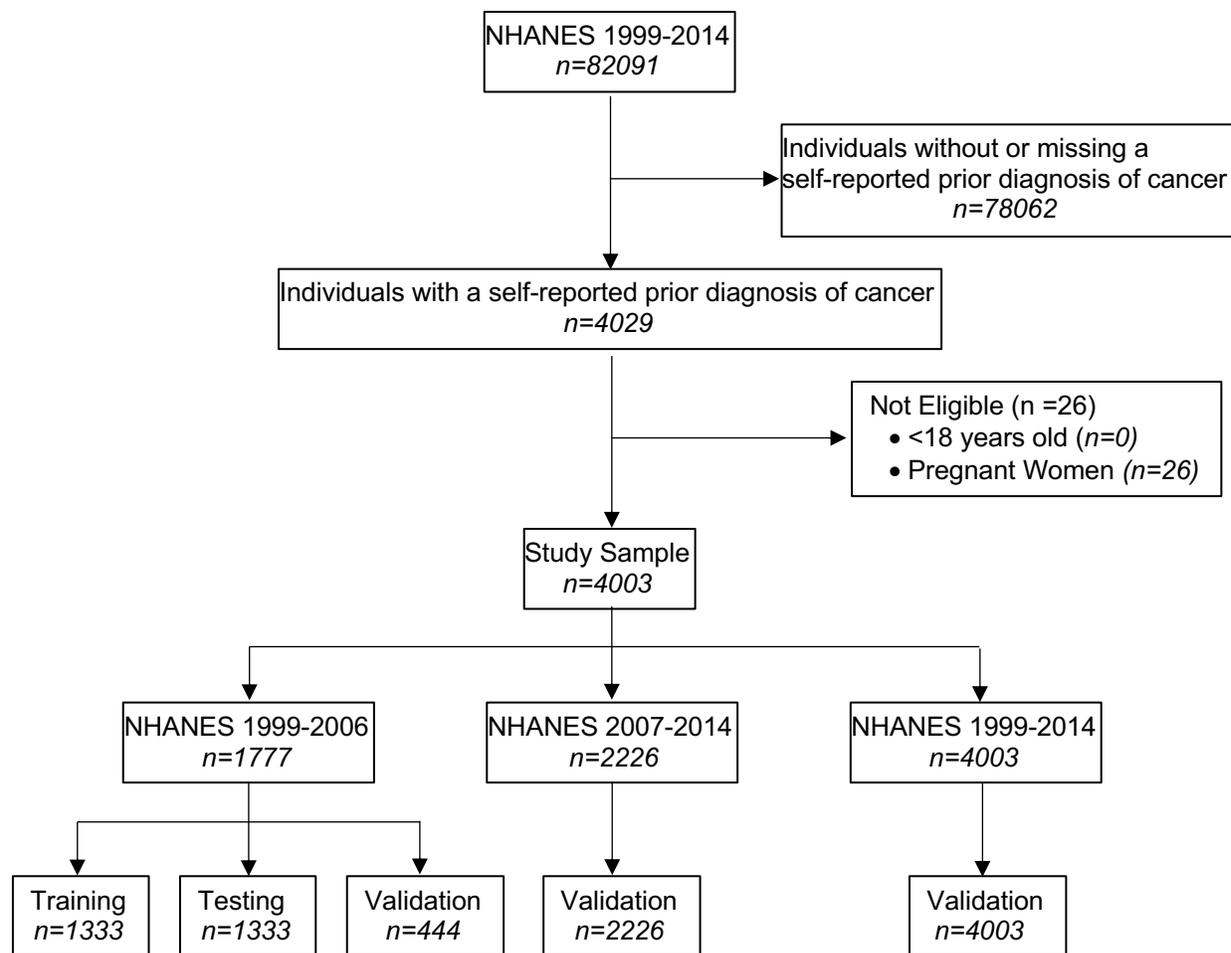
1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30.
2. Group. USCSW. U.S. Cancer Statistics Data Visualizations Tool, based on November 2017 submission data (1999-2015): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. [www.cdc.gov/cancer/dataviz](http://www.cdc.gov/cancer/dataviz). Published June 2018. Accessed June, 2019.
3. Matsuda S, Tsubosa Y, Sato H, et al. Comparison of neoadjuvant chemotherapy versus upfront surgery with or without chemotherapy for patients with clinical stage III esophageal squamous cell carcinoma. *Dis Esophagus.* 2017;30(2):1-8.
4. Kashiwagi S, Yashiro M, Takashima T, et al. Advantages of adjuvant chemotherapy for patients with triple-negative breast cancer at Stage II: usefulness of prognostic markers E-cadherin and Ki67. *Breast Cancer Res.* 2011;13(6):R122.
5. Gianni AM, Piccart MJ. Optimising chemotherapy dose density and dose intensity. new strategies to improve outcomes in adjuvant therapy for breast cancer. *Eur J Cancer.* 2000;36 Suppl 1:S1-3.
6. Prado CM, Baracos VE, McCargar LJ, et al. Body composition as an independent determinant of 5-fluorouracil-based chemotherapy toxicity. *Clin Cancer Res.* 2007;13(11):3264-3268.
7. Jansman FG, Sleijfer DT, Coenen JL, De Graaf JC, Brouwers JR. Risk factors determining chemotherapeutic toxicity in patients with advanced colorectal cancer. *Drug Saf.* 2000;23(4):255-278.
8. Roubenoff R, Kehayias JJ. The meaning and measurement of lean body mass. *Nutr Rev.* 1991;49(6):163-175.
9. Ali R, Baracos VE, Sawyer MB, et al. Lean body mass as an independent determinant of dose-limiting toxicity and neuropathy in patients with colon cancer treated with FOLFOX regimens. *Cancer Med.* 2016;5(4):607-616.
10. Prado CM, Baracos VE, McCargar LJ, et al. Sarcopenia as a determinant of chemotherapy toxicity and time to tumor progression in metastatic breast cancer patients receiving capecitabine treatment. *Clin Cancer Res.* 2009;15(8):2920-2926.
11. Prado CM, Lieffers JR, McCargar LJ, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. *Lancet Oncol.* 2008;9(7):629-635.
12. Morgan DJ, Bray KM. Lean body mass as a predictor of drug dosage. Implications for drug therapy. *Clin Pharmacokinet.* 1994;26(4):292-307.
13. Prado CM, Siervo M, Mire E, et al. A population-based approach to define body-composition phenotypes. *Am J Clin Nutr.* 2014;99(6):1369-1377.
14. Flegal KM, Shepherd JA, Looker AC, et al. Comparisons of percentage body fat, body mass index, waist circumference, and waist-stature ratio in adults. *Am J Clin Nutr.* 2009;89(2):500-508.

15. Kouno T, Katsumata N, Mukai H, Ando M, Watanabe T. Standardization of the body surface area (BSA) formula to calculate the dose of anticancer agents in Japan. *Jpn J Clin Oncol*. 2003;33(6):309-313.
16. Felici A, Verweij J, Sparreboom A. Dosing strategies for anticancer drugs: the good, the bad and body-surface area. *Eur J Cancer*. 2002;38(13):1677-1684.
17. Baker SD, Verweij J, Rowinsky EK, et al. Role of body surface area in dosing of investigational anticancer agents in adults, 1991-2001. *J Natl Cancer Inst*. 2002;94(24):1883-1888.
18. Mourtzakis M, Prado CM, 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(5):997-1006.
19. Laskey MA. Dual-energy X-ray absorptiometry and body composition. *Nutrition*. 1996;12(1):45-51.
20. Bea JW, Zhao Q, Cauley JA, et al. Effect of hormone therapy on lean body mass, falls, and fractures: 6-year results from the Women's Health Initiative hormone trials. *Menopause*. 2011;18(1):44-52.
21. Kilgour RD, Cardiff K, Rosenthal L, Lucar E, Trutschnigg B, Vigano A. Use of prediction equations to determine the accuracy of whole-body fat and fat-free mass and appendicular skeletal muscle mass measurements from a single abdominal image using computed tomography in advanced cancer patients. *Appl Physiol Nutr Metab*. 2016;41(1):70-75.
22. Hinton BJ, Fan B, Ng BK, Shepherd JA. Dual energy X-ray absorptiometry body composition reference values of limbs and trunk from NHANES 1999-2004 with additional visualization methods. *PLoS One*. 2017;12(3):e0174180.
23. Lee DH, Keum N, Hu FB, et al. Development and validation of anthropometric prediction equations for lean body mass, fat mass and percent fat in adults using the National Health and Nutrition Examination Survey (NHANES) 1999-2006. *Br J Nutr*. 2017;118(10):858-866.
24. Yu S, Visvanathan T, Field J, et al. Lean body mass: the development and validation of prediction equations in healthy adults. *BMC Pharmacol Toxicol*. 2013;14:53.
25. National Health and Nutrition Examination Survey Overview. (n.d.). Retrieved from [https://www.cdc.gov/nchs/data/nhanes/nhanes\\_13\\_14/2013-14\\_overview\\_brochure.pdf](https://www.cdc.gov/nchs/data/nhanes/nhanes_13_14/2013-14_overview_brochure.pdf)
26. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National health and nutrition examination survey: sample design, 2011-2014. *Vital Health Stat 2*. 2014(162):1-33.
27. Curtin LR, Mohadjer LK, Dohrmann SM, et al. National Health and Nutrition Examination Survey: sample design, 2007-2010. *Vital Health Stat 2*. 2013(160):1-23.
28. Curtin LR, Mohadjer LK, Dohrmann SM, et al. The National Health and Nutrition Examination Survey: Sample Design, 1999-2006. *Vital Health Stat 2*. 2012(155):1-39.
29. Institute of Medicine (US) Committee on Nutritional Status During Pregnancy and Lactation. Nutrition During Pregnancy: Part I Weight Gain: Part II Nutrient

- Supplements. Washington (DC): National Academies Press (US); 1990. 6, Body Composition Changes During Pregnancy. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK235244/>
30. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination. Examination Protocol. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2005-2006  
<https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Examination&CycleBeginYear=2013>.
  31. Kelly TL, Wilson KE, Heymsfield SB. Dual energy X-Ray absorptiometry body composition reference values from NHANES. *PLoS One*. 2009;4(9):e7038.
  32. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Technical Documentation for the 1999-2004 Dual Energy X-Ray Absorptiometry (DXA) Multiple Imputation Data files. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2008  
[https://wwwn.cdc.gov/Nchs/data/nhanes/dxa/dxa\\_techdoc.pdf](https://wwwn.cdc.gov/Nchs/data/nhanes/dxa/dxa_techdoc.pdf)
  33. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination. Examination Protocol. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2005-2006  
<https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Examination&CycleBeginYear=2013>.
  34. Heymsfield SB, Arteaga C, McManus C, Smith J, Moffitt S. Measurement of muscle mass in humans: validity of the 24-hour urinary creatinine method. *Am J Clin Nutr*. 1983;37(3):478-494.
  35. Kim SW, Jung HW, Kim CH, Kim KI, Chin HJ, Lee H. A New Equation to Estimate Muscle Mass from Creatinine and Cystatin C. *PLoS One*. 2016;11(2):e0148495.
  36. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Laboratory Protocol. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2013-2014
  37. Heeringa S, West BT, Berglund PA. [Applied survey data analysis](#). Second edition. ed. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2017.
  38. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *2011*. 2011;45(3):67.
  39. Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *Am J Epidemiol*. 2014;179(6):764-774.
  40. Liu H, Xu X, Li JJ. A Bootstrap Lasso + Partial Ridge Method to Construct Confidence Intervals for Parameters in High-dimensional Sparse Linear Models. 2017:arXiv:1706.02150.  
<https://ui.adsabs.harvard.edu/abs/2017arXiv170602150L>. Accessed June 01, 2017.

41. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Statistics in Medicine*. 2018;37(14):2252-2266.
42. Harris TB, Visser M, Everhart J, et al. Waist circumference and sagittal diameter reflect total body fat better than visceral fat in older men and women. The Health, Aging and Body Composition Study. *Ann N Y Acad Sci*. 2000;904:462-473.

**Figure 1.** Flow chart describing the development, training, and validation samples for analysis from 8 NHANES cycles 1999-2014



**Table 1.** Distribution of baseline characteristics between men and women with at least one self-reported prior diagnosis of cancer, NHANES Survey Cycles 1999-2006

	Total (n=1777)	Female (n= 933)	Male (n=844)
<b>Demographics</b>			
Age, years (mean, SD)	66.96 (13.78)	64.07 (15.14)	70.17 (11.27)
Race/Ethnicity, n (%)			
Black (non-Hispanic)	218 (12.27)	112 (12.00)	106 (12.56)
White (non-Hispanic)	1364 (76.76)	691 (74.06)	673 (79.74)
Mexican American or Hispanic	162 (9.12)	111 (11.90)	51 (6.04)
Other	33 (1.86)	19 (2.04)	14 (1.66)
Missing/Unknown	0	0	0
Cancer Site, n (%)			
Colon	136 (6.90)	63 (6.12)	73 (7.75)
Rectum	17 (0.86)	6 (0.58)	11 (1.17)
Skin	545 (27.65)	236 (22.93)	309 (32.80)
Lung	60 (3.04)	24 (2.33)	36 (3.82)
Stomach	17 (0.86)	8 (0.78)	9 (0.96)
Pancreas	0 (0.00)	0 (0.00)	0 (0)
Brain	8 (0.41)	3 (0.29)	5 (0.53)
Neck	28 (1.42)	3 (0.29)	25 (2.65)
Blood	4 (0.20)	1 (0.10)	3 (0.32)
Liver	15 (0.76)	8 (0.78)	7 (0.74)
Prostate	298 (15.12)	0 (0.00)	298 (31.63)
Cervix	119 (6.04)	119 (11.56)	0 (0)
Breast	281 (14.26)	281 (27.31)	0 (0)
Other/Unknown	443 (22.48)	277 (26.92)	166 (17.62)
Missing	14	6	8
Age when cancer first diagnosed (mean, SD)	56.70 (17.63)	52.20(18.49)	61.67 (15.16)
Years between most recent cancer diagnosis and DXA measurement, n (%)			
< 1	239 (13.63)	106 (11.52)	133 (15.97)
1	157 (8.96)	71 (7.72)	86 (10.32)
2	114 (6.50)	45 (4.89)	69 (8.28)
3	131 (7.47)	60 (6.52)	71 (8.52)
≥ 4	1112 (63.43)	638 (69.35)	474 (56.90)
Missing	24	13	11
Taken medication(s) for cancer, n (%)	41 (2.31)	35 (3.75)	6 (0.71)
<b>Anthropometric Measures</b>			
Weight, kg (mean, SD)	78.21 (18.45)	73.06 (18.50)	83.83 (16.68)
Height, cm (mean, SD)	167.17 (9.95)	160.65 (7.28)	174.28 (7.23)
Body Mass Index, kg/m <sup>2</sup> (mean, SD)	27.91 (5.89)	28.21 (6.73)	27.58 (4.78)
Body Surface Area, m <sup>2</sup> (mean, SD)	1.90 (0.25)	1.79 (0.24)	2.01 (0.22)
Waist Circumference, cm (mean, SD)	99.19 (14.40)	95.75 (14.88)	102.95 (12.86)
Arm Circumference, cm (mean, SD)	32.00 (4.85)	31.66 (5.41)	32.37 (4.12)
Calf Circumference, cm (mean, SD)	37.59 (4.22)	37.48 (4.60)	37.71 (3.77)
Thigh Circumference, cm (mean, SD)	50.94 (6.91)	51.41 (7.79)	50.44 (5.79)
Subscapular Skinfold, mm (mean, SD)	19.25 (7.61)	19.94 (8.41)	18.52 (6.61)
Triceps Skinfold, mm (mean, SD)	19.14 (8.07)	23.11 (7.56)	15.19 (6.46)
LBM, kg (mean, SD)	48.00 (11.23)	41.11 (7.44)	55.91 (9.51)
LBM % (mean, SD)	61.41 (7.66)	56.64 (5.98)	66.87 (5.38)
Fat Mass, kg (mean, SD)	29.10 (10.88)	31.41 (11.7)	26.44 (9.18)
Fat Mass % (mean, SD)	36.15 (7.96)	41.08 (6.22)	30.48 (5.62)
<b>Physiologic Measures/Biomarkers</b>			
Albumin, ug/mL (mean, SD)	70.39 (395.56)	52.69 (290.29)	89.81 (484.97)
Creatinine, mg/dL (mean, SD)	111.69 (71.66)	94.98 (66.56)	130.03 (72.60)

**Table 2.** Measures of model fit for primary anthropometric prediction equations for total LBM.

	Model 1*		Model 2 <sup>§</sup>		Model 3 <sup>+</sup>		Model 4 <sup>†</sup>		Model 5 <sup>‡</sup>	
	RMSE <sup>++</sup>	R <sup>2††</sup>	RMSE <sup>++</sup>	R <sup>2††</sup>	RMSE <sup>++</sup>	R <sup>2††</sup>	RMSE <sup>++</sup>	R <sup>2††</sup>	RMSE <sup>++</sup>	R <sup>2††</sup>
All	4.26	0.86	2.99	0.93	2.81	0.94	2.81	0.94	2.86	0.94
All Females	2.64	0.93	2.59	0.93	2.61	0.93	2.61	0.93	2.61	0.93
Black	3.78	0.77	3.84	0.76	4.31	0.69			3.85	0.76
White	2.54	0.93	2.49	0.93	2.55	0.93			2.52	0.93
Mexican-American or Hispanic	2.26	0.95	2.57	0.93	2.90	0.92			2.56	0.93
Other	3.66	0.83	3.95	0.81	4.96	0.67			3.81	0.83
All Males	3.39	0.94	3.13	0.95	3.20	0.95	3.20	0.95	3.23	0.95
Black	3.59	0.94	3.21	0.95	3.81	0.93			3.2	0.95
White	3.44	0.94	3.17	0.95	3.21	0.95			3.24	0.95
Mexican-American or Hispanic	3.72	0.88	3.63	0.89	4.61	0.82			3.45	0.90
Other	3.95	0.83	4.43	0.79	8.58	0.24			4.00	0.80

\* Model 1 predictors are height (cm) and weight (kg) only.

§ Model 2 predictors are height (cm), weight (kg) and age.

+ Model 3 predictors are Model 2 predictors plus time interval between cancer diagnosis and DXA scan, sex, race/ethnicity.

† Model 4 predictors consist of Model 3 predictors plus additional interaction terms between anthropometric variables and race/ethnicity.

‡ Model 5 predictors are Model 4 predictors plus all pairwise interactions between anthropometric variables, age, time interval between cancer diagnosis and DXA scan, sex, and race/ethnicity.

++ RMSE = Root Mean Square Error

†† R<sup>2</sup> = R-squared

**Table 3.** Regression coefficients for the most parsimonious models overall, by sex, and by sex and race/ethnicity of the associations between anthropometric measures and total LBM.

Categories	Algorithms
All	$LBM = -67.74 + 0.36^{\ddagger} * (Weight) + 0.52^{\dagger} * (Height)$
All Females	$LBM = -10.64 + 0.32^{\ddagger} * (Weight) + 0.17 * (Height)$
Black	$LBM = 1.23 + 0.31 * (Weight) + 0.12 * (Height)$
White	$LBM = -9.37 + 0.32^{\ddagger} * (Weight) + 0.17 * (Height)$
Mexican-American or Hispanic	$LBM = -13.25 + 0.34^{\S} * (Weight) + 0.18 * (Height)$
Other	$LBM = -37.29 + 0.35^{\dagger} * (Weight) + 0.33 * (Height)$
All Males	$LBM = -33.46 + 0.41^{\ddagger} * (Weight) + 0.32 * (Height)$
Black	$LBM = -31.05 + 0.46^{\ddagger} * (Weight) + 0.29 * (Height)$
White	$LBM = -33.45 + 0.41^{\ddagger} * (Weight) + 0.32 * (Height)$
Mexican-American or Hispanic	$LBM = -34.38 + 0.43^{\dagger} * (Weight) + 0.32 * (Height)$
Other	$LBM = -58.82 + 0.36^{\dagger} * (Weight) + 0.49 * (Height)$

$\ddagger$  P-value  $\leq 0.001$

$\S$  P-value  $\leq 0.01$

$\dagger$  P-value  $\leq 0.05$

**Figure 3.** Correlations between serum creatinine levels (mg/dL) with DXA-based LBM (kg) and derived LBM-algorithm (kg) in the overall sample and by sex and race/ethnicity subgroups.

