

RESEARCH ARTICLE

SUV_{fdg}: A standard-uptake-value (SUV) body habitus normalizer specific to fluorodeoxyglucose (FDG) in humans

Bradley J. Beattie^{1*}, Tim J. Akhurst², Finn Augensen¹, John L. Humm¹

1 Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, United States of America, **2** Division of Radiation Oncology and Cancer Imaging, The Peter MacCallum Cancer Centre, Melbourne, Australia

* beattieb@mskcc.org

Abstract

Purpose

To devise a new body-habitus normalizer to be used in the calculation of an SUV that is specific to the PET tracer ¹⁸F-FDG.

Methods

A cohort of 481-patients was selected for analysis of ¹⁸F-FDG uptake into tissues unaffected by their disease. Among these, 65-patients had only brain concentrations measured and the remaining 416 were randomly divided into an 86-patient test set and a 330-patient training set. Within the test set, normal liver, spleen and blood measures were made. In the training set, only normal liver concentrations were measured. Using data from the training set, a simple polynomial function of height and weight was selected and optimized in a fitting procedure to predict each patient's mean liver %ID/ml. This function, when used as a normalizer, defines a new SUV metric (SUV_{fdg}) which we compared to SUV metrics normalized by body weight (SUV_{bw}), lean-body mass (SUV_{lbm}) and body surface-area (SUV_{bsa}) in a five-fold cross-validation. SUV_{fdg} was also evaluated in the independent brain-only and whole-body test sets.

Results

For patients of all sizes including pediatric patients, the normal range of liver ¹⁸F-FDG uptake at 60 minutes post injection in units of SUV_{fdg} is 1.0 ± 0.16 . Liver, blood, and spleen SUV_{fdg} in all comparisons had lower coefficients of variation compared to SUV_{bw}, SUV_{lbm} and SUV_{bsa}. Blood had a mean SUV_{fdg} of 0.8 ± 0.11 and showed no correlation with age, height, or weight. Brain SUV_{fdg} measures were significantly higher ($P < 0.01$) in pediatric patients (4.7 ± 0.9) compared to adults (3.1 ± 0.6).

Conclusion

A new SUV metric, SUV_{fdg}, is proposed. It is hoped that SUV_{fdg} will prove to be better at classifying tumor lesions compared to SUV metrics in current use. Other tracers may benefit from similarly tracer-specific body habitus normalizers.

OPEN ACCESS

Citation: Beattie BJ, Akhurst TJ, Augensen F, Humm JL (2022) SUV_{fdg}: A standard-uptake-value (SUV) body habitus normalizer specific to fluorodeoxyglucose (FDG) in humans. PLoS ONE 17(4): e0266704. <https://doi.org/10.1371/journal.pone.0266704>

Editor: Val J Lowe, Mayo Clinic, UNITED STATES

Received: November 22, 2021

Accepted: March 25, 2022

Published: April 21, 2022

Copyright: © 2022 Beattie et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available on Dryad <https://doi.org/10.5061/dryad.hqbzkh1j2>.

Funding: This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748. No additional external funding was received for this study.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Standard clinical Positron Emission Tomography (PET) systems typically measure mean radioactivity concentration with a consistency on the order of about 2.5%, this limited primarily by the PET calibration process and the stability of the camera over time [1]. However, radioactivity concentration per se is often *not* a useful metric owing to its variation with the radioactivity of the injected dose. In order to monitor a tumor's uptake of ^{18}F -FDG over several weeks or months, for example, it is necessary to normalize the PET radioactivity concentration by the dose injected at each session, converting it into units of percent injected dose per milliliter (%ID/ml). Meaningful use of this metric assumes a degree of stability of the patient's bodily systems between measurements, consistent timing of the measurement post injection, and linearity of the tissue uptake with injected dose within the range of doses administered (i.e. doubling the injected dose, doubles the tissue concentrations).

While %ID/ml is useful for intra-subject comparisons, it does not allow for meaningful comparisons between patients because it does not account for the variation in tissue uptake of ^{18}F -FDG as a function of the patient body habitus. Larger patients tend to have lower %ID/ml concentrations because the radioactivity is distributed into a larger volume. Thus, to facilitate comparisons of tissue uptake across patients, an additional normalization is necessary. If the radiotracer distribution were to be essentially uniform within the body, then the appropriate additional normalizer would be the patient's body mass (i.e. doubling the patient's size, halves the tissue concentrations). And indeed, this is the normalizer, SUV_{bw} (see Eq 1), that is used most frequently.

$$\text{SUV}_{\text{bw}} = \frac{\text{tissue concentration} \left(\frac{\text{Bq}}{\text{ml}} \right) \left(\frac{\text{ml}}{\text{g}} \right)}{\text{injected dose}(\text{Bq}) / \text{patient weight}(\text{g})} \quad (1)$$

Although the SUV_{bw} metric is, to this day, widely employed, its deficits have frequently been raised, and at no time more strongly than by Keyes who in 1995 [2] concluded that it was a "silly useless value". Most of Keyes' objections could easily be addressed (e.g. by fixing the uptake period) or were not really about the SUV metric itself (e.g. partial volume effects) but at least for ^{18}F -FDG (and likely for many other radiotracers) he correctly pointed out that inter-patient differences in body composition and habitus are not well described by a linear function of body mass alone.

The need for a body habitus normalizer other than body weight stems from the fact that ^{18}F -FDG does not distribute equally into all the normal tissues. On a per unit mass basis, uptake into adipose tissues (in particular), is much less than most other tissues. Thus, two subjects of identical mass but one having a larger fraction of that mass in the form of adipose reserves, will tend to have larger SUV_{bw} values in all their tissues.

Following this reasoning, Zasadny and Wahl in 1993 proposed that FDG uptake be normalized by lean-body mass (SUV_{lbn}) and showed that SUV_{bw} measures of normal blood, liver and spleen all retained a strong correlation with body weight, whereas for SUV_{lbn} , this correlation was greatly diminished. In a similar vein, Kim et al proposed in 1994 [3] normalizing instead by patient body surface-area (SUV_{bsa}) and likewise showed reductions in liver correlation with weight. In neither of these studies was the patient body habitus normalizer (i.e. the actual lean-body mass or body surface-area) measured directly. Instead the normalizer was estimated using simple functions of height and weight, with the lean-body mass estimate making use of two separate functions, one for males and one for females. Kim et al [4] later went on to directly compare SUV_{lbn} to SUV_{bsa} concluding that SUV_{bsa} was superior based upon its relative lack of correlation to body habitus metrics. Nevertheless, in 2009, Wahl et al. [5]

incorporated SUV_{lbm} (a.k.a. SUL) into their PERCIST criteria (the PET equivalent to the CT-based RECIST criteria) proposing it to be used as *the* standard for the evaluation of tumors using ^{18}F -FDG.

Debate over these SUV metrics has continued through to the present day [6], much of this highlighting the vagaries of the lean-body mass and body surface-area estimates [7, 8] each of which can be calculated with one of several different formulas, while others have proposed various means of direct measurement of lean-body mass or body surface area [9–12] or other ancillary corrections [13, 14]. Despite these cogitations and the evidence suggesting that either SUV_{bsa} or SUV_{lbm} would be a better choice, SUV_{bw} remains as the most commonly reported metric in the literature and likely also in clinical use.

In the following, we propose to take a slightly different tact in addressing this question, recognizing that SUV_{bw} , SUV_{lbm} , SUV_{bsa} are all simply functions of patient height, weight and sex, and that maybe none of these surrogates is the optimal body habitus normalizer for ^{18}F -FDG. Based on this premise, we will seek to devise a completely new normalizing function, one that is specific to ^{18}F -FDG. As was the case in the previous evaluations of SUV metrics, we will assume that ^{18}F -FDG uptake in a normal liver does not itself vary systematically with body habitus, age or sex. Moreover, we assert that confounding factors of any sort can only increase an SUV metric's coefficient of variation (CoV) above the liver's true normal range and thus smaller CoV values are indicative of a less biased normalizer.

Materials and methods

Patients

The data used in this study was derived from patients receiving standard of care ^{18}F -FDG scans at our institution, mostly for the diagnosis and monitoring of cancerous lesions. Patients were excluded if they were diagnosed with a non-solid tumor type, had extensive disease, had any indication of lesions within an organ being measured or were imaged outside of the 55–75 minute post injection time window recommended by the European Association of Nuclear Medicine (EANM) [15] and the Quantitative Imaging Biomarkers Alliance (QIBA) [16]. A total of 481 patients meeting these criteria were included in the study. A subset of these (100 in all) were specifically sought after, selected based on their age (15 or under) in order to enrich the sample with smaller sized subjects.

Of the 481 patients, 65 had only their normal brain ^{18}F -FDG uptake measured. The remaining 416 patients were randomly divided into a 330 subject training group that received only normal liver ^{18}F -FDG uptake measurements and an 86-member test group within which normal liver, spleen and blood concentrations were measured. Of the 330 training group members, 153 were adult women, 116 were adult men and 61 were pediatric patients (note—here the division between pediatric and adult was taken to be 12 years of age, i.e. “adults” > 12 y). Within the test cohort there were 45 adult women, 31 adult men and 10 pediatric patients. And within the brain-only cohort, there were 14 adult women, 29 adult men and 22 pediatric patients.

Subjects were included regardless of what PET scanner model was used, so the cohort includes a mixture of scans from various GE PET cameras including Discovery PET/CT models DST, DSTE, D600, D690, D710, 3-ring DMI, 5-ring DMI and a Signa PET/MR. This data was analyzed under the auspices of a retrospective research protocol “Image Processing Applications for Medical Imaging Workstations and Systems”, IRB# 16-1488A(12), which was approved by the Memorial Sloan Kettering Cancer Center Institutional Review Board as Exempt, and the requirement for consent was waived.

Measurements

Within the training and test cohort patient scans a single large region of interest (ROI) representing a volume of approximately 14 ml, was drawn well away from the diaphragm. Within the test cohort scans, additional ROIs were placed over homogenous regions well within the descending aorta (~1 mL) to measure the blood concentration, and spleen (~2.5 mL). Within scans of the brain-only test cohort, a single ROI was placed over a frontal grey matter region (~0.5 mL). In all cases the ROI was drawn free-hand on a single slice encompassing over a region that was homogeneous in its FDG uptake and (based on visualization of the entire tissue) representative of the tissue as a whole. All were drawn on transaxial images except for the descending aorta region which was drawn on a sagittal slice extending vertically along the aorta's length and referencing the co-acquired CT. In addition to the mean radioactivity concentration within these regions, the following measures describing the patient scan were compiled: patient age at time of scan, weight, height, sex, injected radioactivity and the time interval between the injection and when the bed position over a measured region was acquired. All radioactivity concentration measures were appropriately decay corrected and divided by the injected activity to arrive at units of %ID/ml. This value was then multiplied by the patient's body weight in grams, which if one assumes 1 g/ml, results in unitless SUV_{bw} values. The values were also multiplied by the calculated body surface-area and lean-body mass to arrive at SUV_{bsa} and SUV_{lbm} measures, respectively; making use of the body-surface area estimation function proposed by Du Bois [17] and the lean-body mass function used by Lodge and Wahl for PERCIST [18].

Model development

In seeking an empirical functional form that would well describe the relationship between the liver mean %ID/ml and body habitus, we first reasoned that these two quantities should be roughly inversely proportional and therefore chose to attempt to model the multiplicative inverse of the liver %ID/ml (i.e. its mean concentration in units of ml/%ID). Moreover, since it was our preference that our model achieve specifically a high *percent* accuracy and result in only positive normalizing values, we chose to fit its log values (i.e. $\log[\text{ml}/\%ID]$).

Through some experimentation with the training set, least squares fits of various functions were compared [Curve Fitting Toolbox v 3.5.11, The MathWorks, Inc.] and a subjective "best" was selected making use of Bayesian (BIC) [19] and Akaike information criteria (AIC) [20], the adjusted R-squared value [21] of the fits and a visual examination of the residuals.

Model validation and testing

Using the selected fitting function model, the training set was then entered into a 5-fold cross-validation study. In this study the training set was first randomly divided into five subgroups each containing 20% (i.e. 66) of the patients. Each of the 5 groups was then, in turn, used as a validation set, with the remaining 80% (264 patients) used to train (i.e. fit) the model. In each of the five validations, CoVs and correlations to height and weight for each of the four SUV metrics (SUV_{bw} , SUV_{lbm} , SUV_{bsa} and SUV_{fdg}) were calculated and based on these numbers the performance of our proposed body habitus normalizing (BHN) function was assessed.

Following this validation, a single fitting procedure using the selected BHN model was applied to the entire 330 patient training set to determine its parameter values. This BHN function was then used to calculate the SUV_{fdg} values for all the normal tissue measurements taken from the two test sets. As was done in the cross-validation, SUV_{bw} , SUV_{lbm} and SUV_{bsa} values were determined and compared based on their CoVs and correlations to height and weight, but in addition, for the test cohorts, correlation to age was also tested.

Statistics

For every test of a linear relationship between a variable (SUV, residual, etc.) to patient height, weight or age, a Pearson's correlation coefficient R and associated P value were determined. This P value indicates the probability of seeing a sample correlation coefficient of that magnitude when the true population correlation is zero and was calculated using two tails of a t -distribution with $n-2$ degrees of freedom (where n is the number of samples) after first converting the R value to a t -statistic using the formula $t = \frac{R\sqrt{n-2}}{1-R^2}$. In all cases significance was assessed at an alpha level of 0.05, corrected for multiple comparisons following Bonferroni [22] where indicated. The comparison of adult and pediatric brain SUV_{fdg} values was made with an unpaired two-tailed, two sample t -test assuming unequal variances.

Results

Cohort characterization

Subjects ranged from 9 months to 91 years of age and were roughly evenly distributed over this range (see S1 Fig) owing to the enriched selection of pediatric patients. Women tended to be smaller in both their height, 1.62 ± 0.07 m, and weight 70 ± 19 kg, compared to the men who tallied in at 1.74 ± 0.09 m and 83 ± 20 kg, respectively. Pediatric patients (under 12) averaged 1.17 ± 0.20 m in height and weighed 23 ± 11 kg.

Model development

Using data from the training set only, visual inspection of the weight vs. $\log(\text{ml}/\%ID)$ and height vs. $\log(\text{ml}/\%ID)$ suggested that simple functions of weight did not fit the data well (see S2 Fig) whereas a third order polynomial function of height estimated the $\log(\text{ml}/\%ID)$ values in an unbiased manner (see Fig 1A). As a means of confirming this, second and fourth order polynomial functions of height were also tried and compared based on AIC, BIC and adjusted R^2 values (see Table 1). The AIC and adjusted R^2 values showed a slight preference for the 3rd order model, but the BIC was best for the 2nd order polynomial function, therefore both models remained under consideration.

Although functions of weight alone did not appear to predict the liver concentrations well, there was still a potential that the addition of height information might improve the fit substantially. Therefore, we added a linear term incorporating height to the 3rd order function of weight (see Model C in Eq 2). However, the fit continued to be poor, especially for small patients (see S3 Fig) and so we dropped Model C from further consideration.

Then to ascertain whether adding weight information might improve the estimate of the 3rd order function of height model, we plotted its residuals as a function of patient weight (see Fig 1B). This plot showed that there was remaining correlation which could perhaps be improved if weight were to be incorporated. This potential also remained for the 2nd order function of height, so to each of these models was added a single parameter, d , incorporating the weight information. We will hereafter refer to the 3rd order height plus 1st order weight function as Model A and the 2nd order height plus 1st order weight as Model B (see Eq 2). Adding weight information in this way to Model A removed all correlation of the residuals with either height or weight (see Fig 1C and 1D, respectively).

$$\log[\text{ml}/\%ID] = ax^3 + bx^2 + cx + dy + e$$

where for model A : $x = \text{height and } y = \text{weight}$

for model C : $x = \text{weight and } y = \text{height}$

for model B : $a = 0, x = \text{height and } y = \text{weight}$

(2)

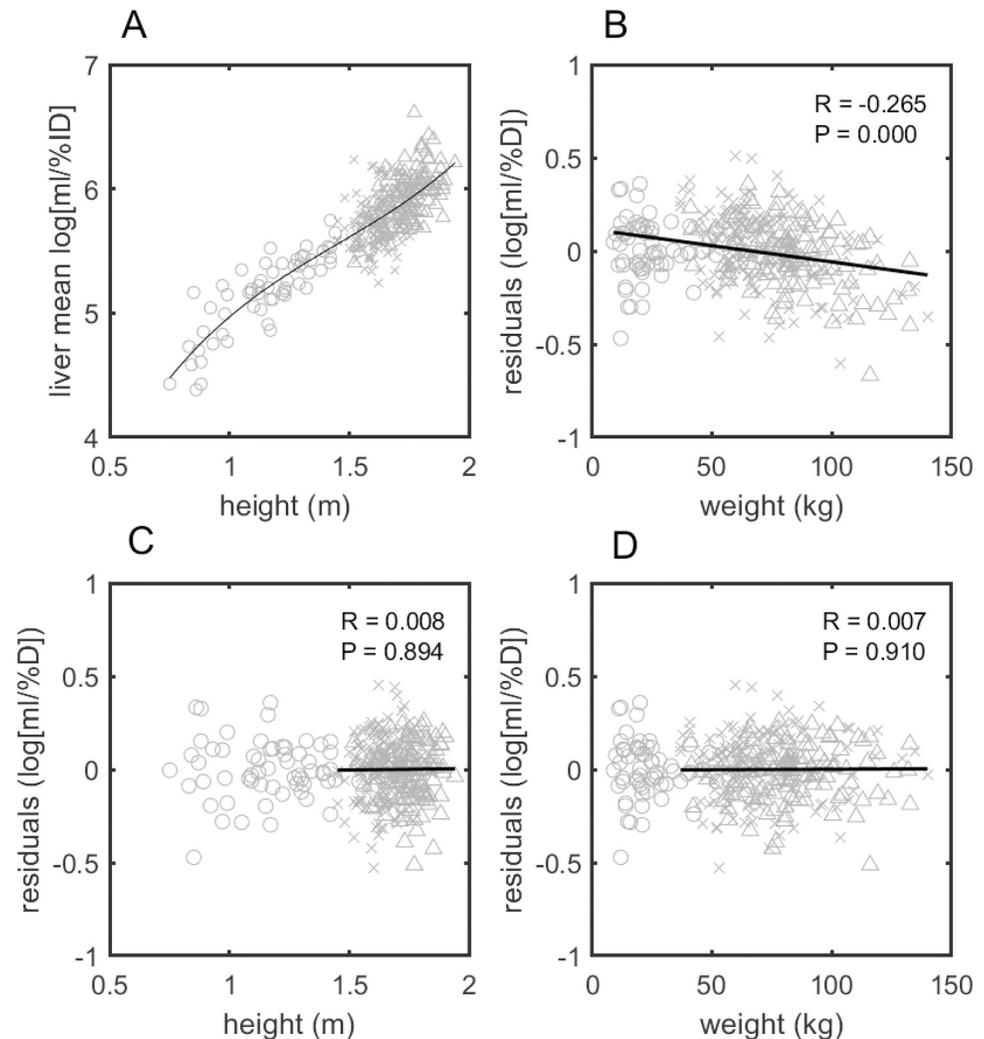


Fig 1. Training data, liver mean fits and residuals as a function of height and weight. Scatter plots showing the fit of the training set data to a third-order polynomial function of height (A), the residuals of that fit as a function of weight (B), the residuals of the model A fit to the log of the mean liver ml/%ID as a function of height (C) and its residuals as a function of weight (D). In all graphs, triangles depict male patients, x's refer to female patients and o's are children under the age of 12. The fits of the residuals shown in (C) and (D) along with the associated correlation coefficients and *P* values shown in the legend, excluded patients under the age of 12 so that these patients wouldn't have outsized influence over the correlation. For Model A, regardless of whether the pediatric patients were included or not, there was no significant correlation of the residuals to either height or weight.

<https://doi.org/10.1371/journal.pone.0266704.g001>

Table 1. Model information criteria evaluation.

	2nd order	3rd order	4th order	Model A	Model B
AIC	-1110	-1111	-1110	-1185	-1176
BIC	-1099	-1096	-1091	-1166	-1161
adjR ²	0.725	0.727	0.726	0.782	0.775

Akaike and Bayes Information Criteria and Adjusted R² values for fits of various models to log(ml/%ID) of the training cohort. The 2nd, 3rd and 4th order models are a function of height alone. Models A and B are functions of height and weight.

<https://doi.org/10.1371/journal.pone.0266704.t001>

Model A and Model B were both fit to the training set data. In this instance, however, the AIC, BIC and adjusted R^2 values (again see Table 1) were all better for model A. The residuals for Model A's fit to the data are plotted as functions of height and weight in Fig 1C and 1D, respectively. In these plots we emphasize that there remains no correlation to body habitus even when restricted to the adult population by showing the fit based only on those patients. The correlation remained near zero, however, when the pediatric patients were included. Based on these assessments we selected Model A as the functional form to be used as the BHN when calculating SUV_{fdg} .

Model testing

Using model A as the functional form for our proposed BHN, a 5-fold cross validation assessment comparing SUV_{fdg} to SUV_{bw} , SUV_{lbm} and SUV_{bsa} was conducted using data in the training set. In this assessment, the data was randomly partitioned into 5 validation groups each containing 66 patients. Then in turn for each of these groups, the remaining 264 patients were used to fit the five parameters of model A. The means and standard deviations for each of the coefficients over the five fits were as follows: $a = 1.46 \pm 0.13$, $b = -6.48 \pm 0.56$, $c = 10.10 \pm 0.74$, $d = 0.00512 \pm 0.000226$, $e = -0.168 \pm 0.312$. The resulting normalizing function was then applied to calculate SUV_{fdg} values (see Eq 3) for the validation group along with values for SUV_{bw} , SUV_{lbm} , and SUV_{bsa} .

$$SUV_{fdg} = \frac{\text{tissue concentration} \left(\frac{Bq}{ml} \right)}{\text{injected dose (Bq)} / BHN(ml)} = \frac{\%ID}{ml} \cdot \frac{BHN(ml)}{100} \quad (3)$$

Coefficients and P values were calculated for each SUV's correlation to height and weight. The P values were each assessed at an alpha level of 0.05 but Bonferroni corrected [22] for the 5 tests (i.e. were considered significant at P values < 0.01). CoVs for each of the SUVs were also determined. The results of this assessment are shown in Table 2. The mean of the five SUV_{fdg} CoVs, 0.16, was taken to be the best estimate of the population normal liver SUV_{fdg} standard deviation and was reported in the Abstract. In all five validations, SUV_{fdg} had the smallest CoV, followed by SUV_{bsa} , then SUV_{lbm} with SUV_{bw} having the largest CoV. This same pattern was seen in the correlation coefficients describing the relationship of the SUV metrics to both patient height and weight, with SUV_{fdg} always showing the lowest correlation. In five out of five tests, the correlation coefficients to height and weight for SUV_{bw} and SUV_{lbm} were significant. In one out of five tests the correlation of SUV_{bsa} to height was significant, but no correlations of SUV_{bsa} to weight occurred. In no instance was the correlation of SUV_{fdg} to either height or weight found to be significant.

Table 2. Cross validation results.

	statistic	SUV_{bw}	SUV_{lbm}	SUV_{bsa}	SUV_{fdg}
weight	CoVs	0.289 ± 0.010	0.223 ± 0.015	0.175 ± 0.014	0.162 ± 0.014
	R-value	0.768 ± 0.051	0.523 ± 0.020	0.217 ± 0.019	-0.000 ± 0.068
	P value	0.000 ± 0.000	0.000 ± 0.000	0.083 ± 0.030	0.695 ± 0.216
	N of 5 sig	5	5	0	0
height	R-value	0.602 ± 0.096	0.623 ± 0.062	0.234 ± 0.064	-0.008 ± 0.096
	P value	0.000 ± 0.000	0.000 ± 0.000	0.082 ± 0.065	0.541 ± 0.194
	N of 5 sig	5	5	1	0

Results from the 5-fold cross validation comparing SUV_{bw} , SUV_{lbm} , SUV_{bsa} and SUV_{fdg} based upon the coefficients of variation (smaller is better) and correlations to height and weight (again, smaller is better). The "N of 5" rows indicate the number of times a correlation coefficient was determined to be significant.

<https://doi.org/10.1371/journal.pone.0266704.t002>

BHN parameter determination

Model A's final parameter values were determined in a single least-squares fit of the entire 330 patient training set and are described in Eq 4, wherein height is measured in meters, weight in kilograms and the coefficients are all taken to have units such that the unit of the final BHN result is in milliliters. Coefficient values are shown, followed by the 95% confidence interval in parenthesis.

$$BHN = 100e^{(ax^3+bx^2+cx+dy+e)}$$

where : $x = \text{height and } y = \text{weight}$

$$a = 2.03 (1.13, 2.93)$$

$$b = -9.07 (-12.92, -5.22) \quad (4)$$

$$c = 13.94 (8.56, 19.33)$$

$$d = 0.00539 (0.00438, 0.00641)$$

and : $e = -2.04(-4.49, 0.41)$

The quality of these fits to this function can be appreciated by viewing the 3D scatter plot showing the proposed BHN function surface on log[ml/%ID] vs. height vs. weight axes (S4 Fig). Moreover, Figs 2A–2D and 3A–3D show the correlations for the full training cohort to weight and height, respectively, of the standard SUV metrics (SUV_{bw} , SUV_{lbm} and SUV_{bsa}) in comparison to that of the SUV_{fdg} metric. The results here essentially recapitulate those of the 5-fold cross validation with SUV_{bw} and SUV_{lbm} both having clearly significant correlation to body habitus, SUV_{bsa} showing weak correlation, and SUV_{fdg} having no significant correlation.

Test cohort results

The training set results for the liver were confirmed in the independent test cohort. Scatter plots show no significant correlation of SUV_{fdg} to either height or weight (see Fig 4), but also, no correlation to patient age, a parameter not considered in the determination of the BHN model. Importantly, these improvements also extended to tissues not at all used in the derivation of the BHN function. Scatter plots showing the correlations with height, weight and age for the normal spleen are shown in S5 Fig and for blood in Fig 5. That SUV_{fdg} appears to be a good predictor of the patient's blood concentration (at this time post injection) is particularly significant given the relationship between the area under the blood time vs. activity curve (TAC) and absolute quantitative uptake of ^{18}F -FDG.

Interestingly, measurements of gray matter uptake taken from the independent brain-only test cohort, show a small reduction in all four SUV metrics as a function of age in adult patients (see S6 Fig). These decreases did not reach statistical significance but are consistent with at least one other study which showed reduced brain glucose metabolic rates in older adults based on a modeled quantitative assessment of ^{18}F -FDG uptake [23]. SUV_{fdg} , SUV_{bsa} and to a lesser extent SUV_{lbm} all showed noticeably higher levels in the pediatric patients (≤ 18 y) within this cohort. The SUV_{fdg} values for these two groups, 4.7 ± 0.9 for pediatric patients compared to 3.1 ± 0.6 for adults, were found to be significantly different ($P < 0.01$) in a two-sample t-test. This was not the case, however, for SUV_{bw} . All CoV and correlation results for all tissues measured in the independent test cohorts were also tabulated (see Table 3).

Discussion

Herein we propose a new body habitus normalizer to be used when calculating SUV values within PET ^{18}F -FDG patient studies. This body habitus metric, like the estimates of lean-body

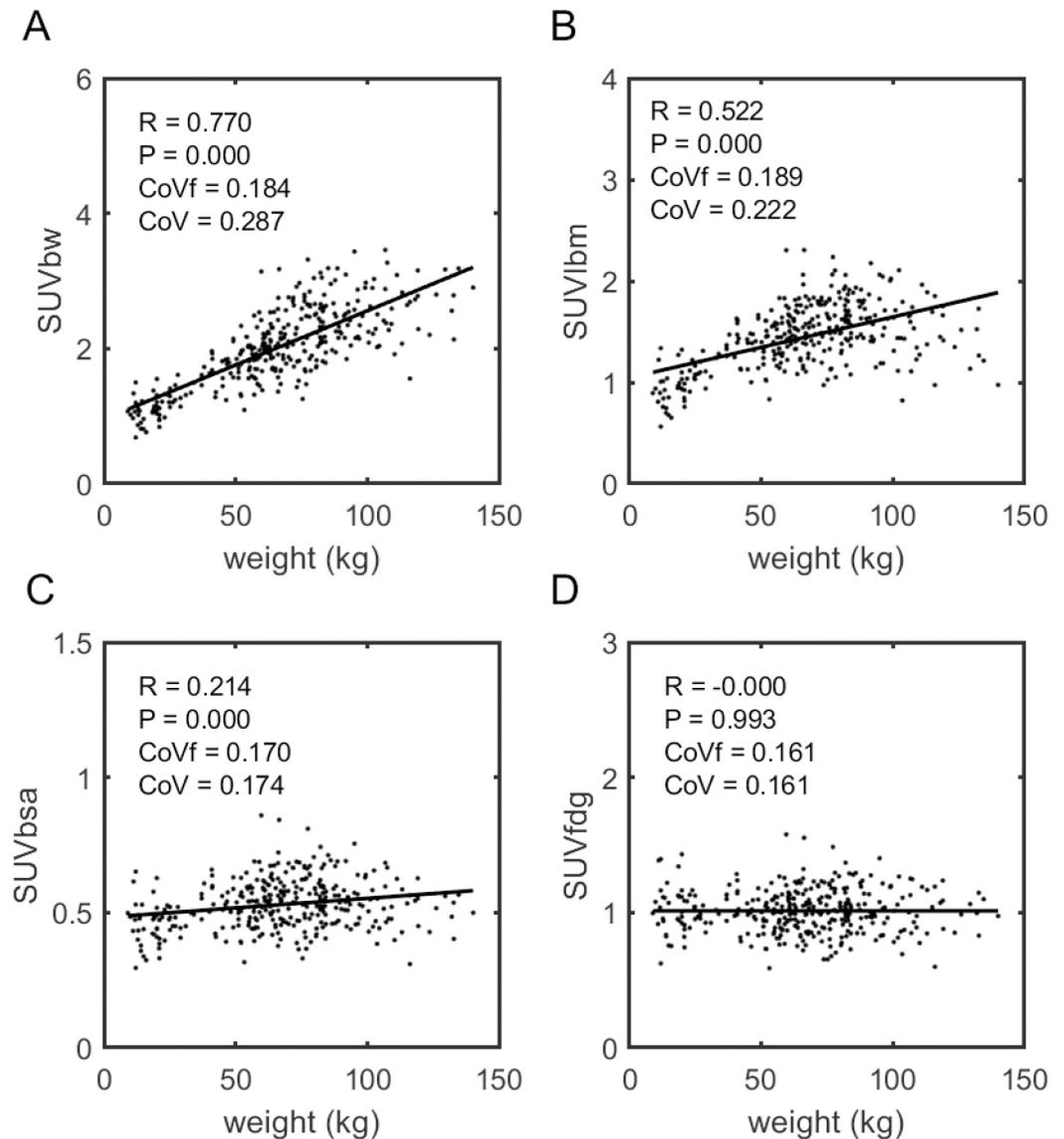


Fig 2. Training data, SUVs as a function of weight. Scatter plots showing the correlation of various types of liver SUV measurement to patient weight within the training cohort, (A) SUV_{bw} , (B) SUV_{lbm} , (C) SUV_{bsa} and (D) the proposed SUV_{fdg} metric. Only SUV_{fdg} shows no significant correlation to weight. The CoVf value describes the variance about the fitted line while CoV describes the variance about the mean.

<https://doi.org/10.1371/journal.pone.0266704.g002>

mass and body-surface area, is based on simple measures (height and weight) of the patient that can be determined prior to imaging. Like SUV normalized by body weight, the SUV_{fdg} metric calculated using the proposed normalizer can be considered unitless. The value can be interpreted as a fraction of the expected normal liver mean uptake at (approximately) 60 minutes post injection wherein normal liver is expected to have a value of 1.0 and therefore any tissue with an SUV_{fdg} value of 2.0, for example, has twice the liver's uptake. As such, this metric can also be used as a quick quality assurance measure to identify data entry errors for the injected dose, its timing, or errors in the entry of the patient height or weight. Values dramatically different from 1.0 ± 0.16 measured in a normal region of a patient's liver would be indicative of a problem, including perhaps significant extravasation of the injectate.

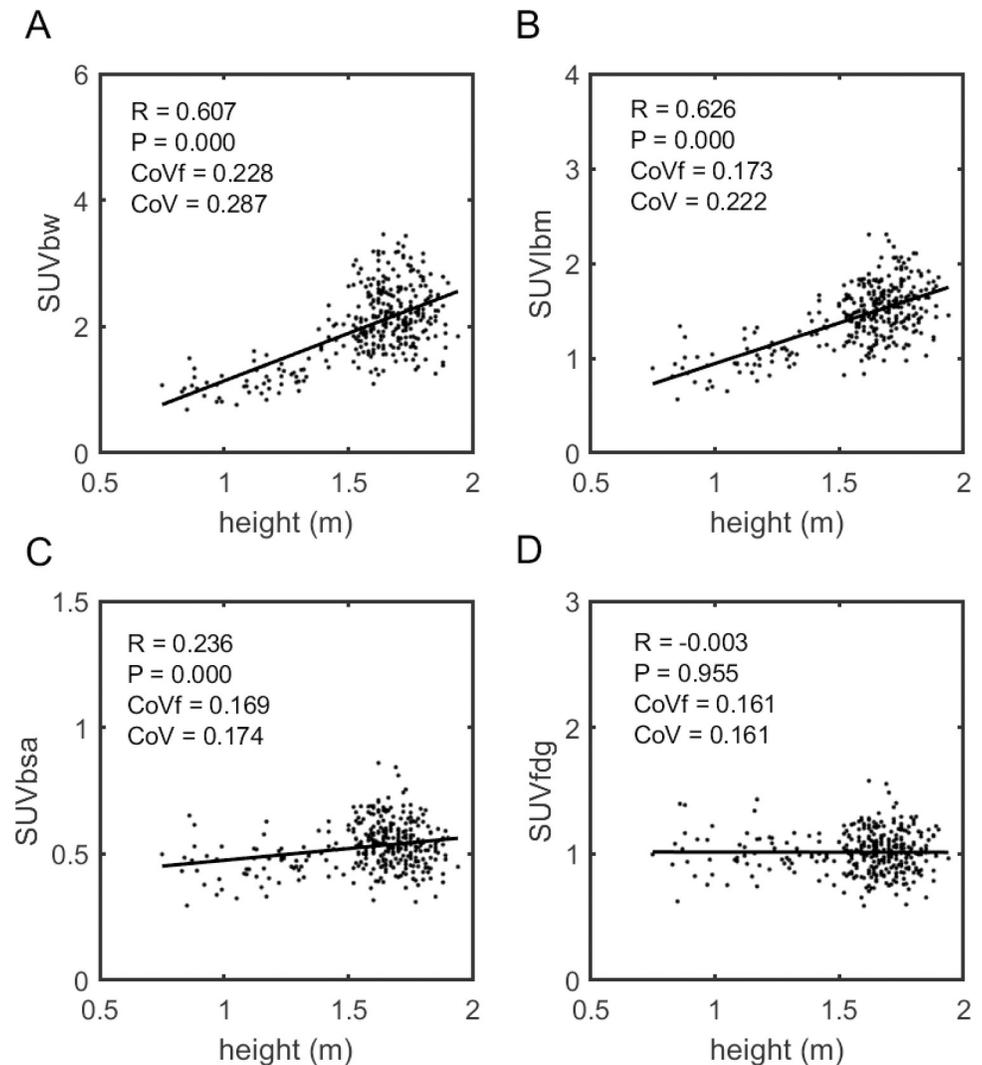


Fig 3. Training data, SUVs as a function of height. Scatter plots of liver measurements like those shown in Fig 2 except here plotted as a function of patient height. Again, only SUV_{fdg} has no significant correlation.

<https://doi.org/10.1371/journal.pone.0266704.g003>

To the extent that the proposed BHN function can accurately predict the uptake to the normal liver and proportionately that of other normal organs (in units of %ID/g), this function may be useful in models seeking to estimate patient-specific radiation dose, thus allowing an a priori individualized assessment of the risk posed by the ¹⁸F-FDG injection. Similarly, this same information can be used in models of patient attenuation and scatter, which can then be combined with models of specific PET cameras to arrive at estimates of the expected noise equivalent count (NEC) rate for different body-parts. This information can then, in turn, be used to adjust imaging time to achieve a target image quality. Assuming the intrinsic resolution of most clinical PET cameras is about the same (or can be made so with appropriate smoothing) matching total effective NECs (factoring in the use of time-of-flight and the camera's timing resolution) should go a long way towards harmonizing image quality across patients of different sizes and across institutions having a mixture of PET camera models.

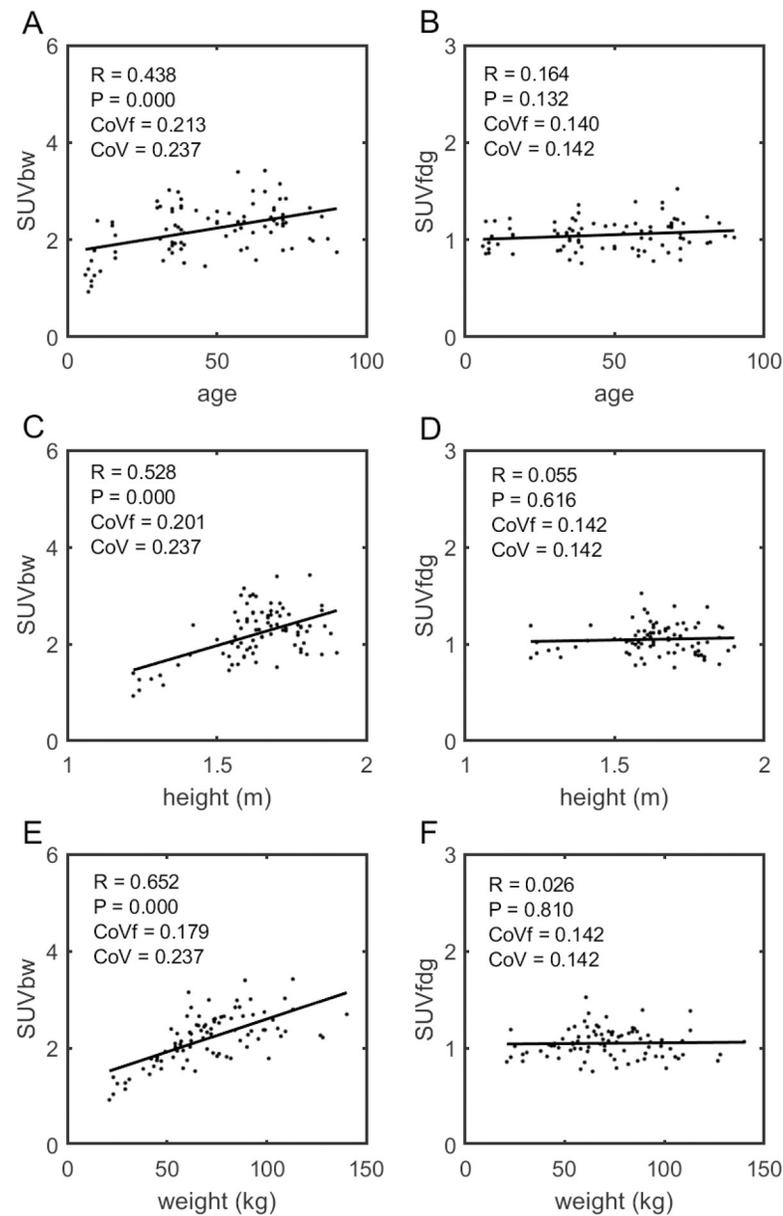


Fig 4. Liver test data, SUVs as a function of age, height and weight. For the independent test data, these scatter plots compare the correlations in liver SUV_{bw} (column A, C, E) and SUV_{fdg} (column B, D, F) measurements with age (row A, B), height (row C, D) and weight (row E, F).

<https://doi.org/10.1371/journal.pone.0266704.g004>

Although SUV_{fdg} is specific to ¹⁸F-FDG PET, the concept behind it should be applicable to all tracers for which a suitable normal reference tissue can be found, and where any metabolism is either consistent or at least predictable across patients.

Limitations

While in principle each radiotracer could/should have its own optimal body-habitus normalizer, this normalizer may not be well represented as a function of simple patient descriptors (height, weight, sex). For example, differences in tracer metabolism or excretion among

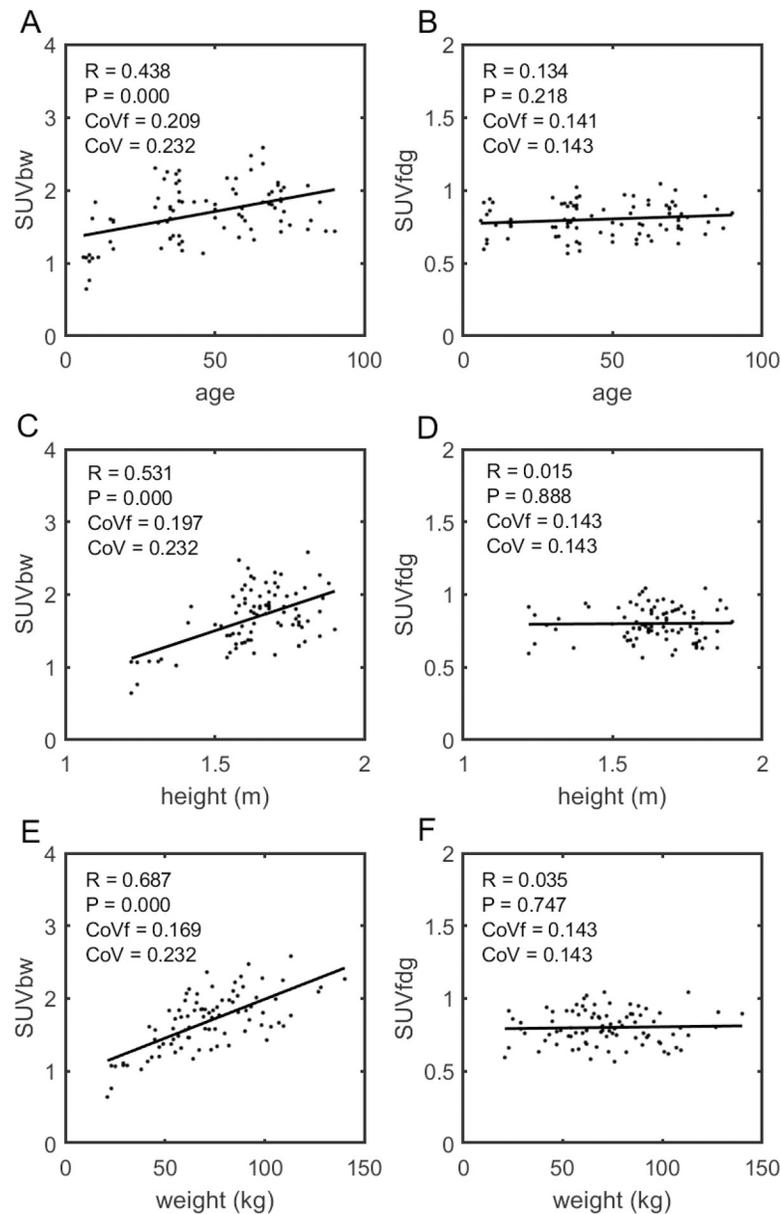


Fig 5. Blood test data, SUVs as a function of age, height and weight. For the independent test data, these scatter plots compare the correlations in blood SUV_{bw} (column A, C, E) and SUV_{fdg} (column B, D, F) measurements with age (row A, B), height (row C, D) and weight (row E, F). Note, blood concentrations were not measured in the training cohort and played no part in determining the BHN function used to calculate these SUV_{fdg} values.

<https://doi.org/10.1371/journal.pone.0266704.g005>

patients could easily be the dominant factor in determining tracer uptake without correlation to body-habitus. And even when some body-habitus metric could work, difficulty finding an appropriate reference normal tissue that is sufficiently large and low noise might hamper the ability to define a good normalizer.

While we have yet to directly demonstrate improved performance with SUV_{fdg} compared to other SUV metrics when applied to clinically relevant questions, we wish to point out that “improved performance” may be difficult to define in that it could mean either finding correlation where none was seen previously, or removing correlation with a clinical metric that was

Table 3. Independent test cohort results.

	statistic	CoV	weight		height		age	
			R	P	R	P	R	P
LIVER	SUV _{bw}	0.24	0.65	0.00	0.53	0.00	0.44	0.00
	SUV _{lbm}	0.18	0.43	0.00	0.54	0.00	0.49	0.00
	SUV _{bsa}	0.17	0.07	0.54	0.13	0.22	0.29	0.01
	SUV _{fdg}	0.14	0.03	0.81	0.05	0.62	0.16	0.13
BLOOD	SUV _{bw}	0.23	0.44	0.00	0.53	0.00	0.69	0.00
	SUV _{lbm}	0.16	0.28	0.01	0.10	0.34	0.07	0.50
	SUV _{bsa}	0.18	0.50	0.00	0.54	0.00	0.47	0.00
	SUV _{fdg}	0.14	0.13	0.22	0.02	0.89	0.04	0.75
SPLEEN	SUV _{bw}	0.23	0.61	0.00	0.41	0.00	0.34	0.00
	SUV _{lbm}	0.16	0.35	0.00	0.38	0.00	0.39	0.00
	SUV _{bsa}	0.16	-0.08	0.44	-0.13	0.23	0.09	0.39
	SUV _{fdg}	0.16	-0.16	0.13	-0.26	0.01	-0.07	0.51
BRAIN	SUV _{bw}	0.21	-0.15	0.41	0.16	0.37	-0.30	0.09
	SUV _{lbm}	0.19	0.13	0.48	0.33	0.06	-0.11	0.53
	SUV _{bsa}	0.19	-0.07	0.69	-0.06	0.73	-0.01	0.97
	SUV _{fdg}	0.19	0.11	0.55	0.16	0.38	-0.11	0.55

Variance and correlation results for SUV_{bw}, SUV_{lbm}, SUV_{bsa} and SUV_{fdg} metrics for data measured from patients within the independent test cohorts. ¹⁸F-FDG concentrations were measured in normal liver, blood, spleen and brain. CoV is the coefficient of variation, R is the correlation coefficient and P is the probability of seeing an R value of that magnitude when the true population correlation is zero.

<https://doi.org/10.1371/journal.pone.0266704.t003>

in truth dependent upon a parameter confounding the original SUV metric. In other words, if for example mean SUV_{bw} of a tumor was found to predict survival, it's conceivable that this prediction was actually driven by patient body weight, largely or completely independent of FDG uptake in the tumor. In such cases, the improved performance of SUV_{fdg} might mean the spurious correlation would no longer be significant. Nevertheless, we will be applying SUV_{fdg} if future studies seeking to use FDG uptake as biomarker to seek correlations or divide tumors into subgroups. We expect that SUV_{fdg} will prove to be particularly useful in patient cohorts involving a wide range of patient sizes including potentially pediatric subjects.

The SUV_{fdg} metric we propose is no doubt imperfect. It is a function of just two parameters (patient height and weight) and was not found using an exhaustive search of potential functional forms. In no way do we claim that it is optimal. This we believe is in keeping with previously published SUV metrics. Normalization to reference tissues, or adjustments based on blood glucose measurements, or normalizations based on direct measurements of fat, muscle, and other normal tissue volumes, may all prove to be better than the metric we propose, in some contexts, however, in keeping with the spirit of SUV-type measurement, the metric we propose is applicable in all contexts, regardless of what body-parts are scanned and regardless of the availability of other refining variables. Moreover, we feel we have demonstrated (hopefully) convincingly that SUV_{fdg} is less confounded and has less variability than SUV_{bw}, SUV_{lbm} and SUV_{bsa}.

A key assumption when calculating and using this new body-habitus normalizer is that the rate constants governing normal liver ¹⁸F-FDG uptake are essentially the same (i.e. within a normal range) across all subjects regardless of age or sex. In other words, we have assumed that normal liver is itself a good normalizer, one whose uptake is proportional to the area under the curve of the arterial blood input function up until the time of the PET measurement at 60 minutes post injection. This assumption is strongly supported by our SUV_{fdg}

measurements of the blood. As such, the proposed function should also be a useful normalizer for most other normal and abnormal tissues within the body.

One potential caveat to this assumption and possible confound in this study, is that we did not screen for fatty liver disease or other liver morbidities that might correlate with body habitus [24]. As such, the proposed BHN in its current form effectively includes estimations of the prevalence and impact of these disease processes in our population. Similarly, if liver ^{18}F -FDG uptake in absolute terms varied significantly with age in the pediatric population, the proposed SUV_{fdg} metric would normalize away that difference given the high correlation between age and height in children. In other words, an SUV_{fdg} value of 1.0 can be considered to be the normal liver uptake level for pediatric subjects regardless of age even if the uptake in units of $\text{mg}/\text{min}/100\text{g}$ were to go up or down as a function of age. Given our results in the blood, however, we feel these effects are at most, small.

When evaluating the performance of SUV_{fdg} relative to the other SUV types we have relied on two related assessments, each SUV metric's CoV and the absence of any correlation to body habitus (specifically height and weight). The assessment based on correlation to body habitus has been used by others [3, 4, 25] but to our knowledge we are the first to make use of CoV for this purpose. In using CoV to assess SUV's, we reasoned that the optimal SUV metric should accurately reflect the normal variation in liver ^{18}F -FDG uptake and that any additional noise or confounds would only lead to increased CoV. This should be true so long as the variance in normal values, which presumably are randomly distributed about the mean, is not itself correlated with the SUV metric. It is anticipated that because of its reduced CoV and correlation to body habitus, it will likely outperform SUV_{bw} , SUV_{lbm} and SUV_{bsa} when used to distinguish between two or more conditions, for example when classifying benign and malignant tumors across multiple patients.

Conclusion

A new body habitus normalizer and associated SUV metric are proposed. This metric, SUV_{fdg} , is intended to be used solely for the evaluation of the uptake of FDG and may in future studies be shown to outperform SUV metrics normalized by body weight, lean-body mass and body surface area.

Supporting information

S1 Fig. Age distribution of males and females in whole cohort. Histograms showing the distribution of ages for the entire cohort of 481 patients. The frequencies for males (A) and for females (B). These plots show that the sampling was roughly uniform with respect to patient age, a result achieved owing to an enhanced search for younger patients.
(PDF)

S2 Fig. Poor fit of liver training data, log inverse concentration as function of weight alone. Scatter plot of training set data showing relationship between patient weight and mean liver concentration expressed in units of $\log(\text{ml}/\%ID)$ and fitted with a 3rd order polynomial. This fit performs poorly for subjects below 20 kg.
(PDF)

S3 Fig. Poor fit of liver training data, log inverse concentration as function of both height and weight. Same data as shown in Figure S2 except now as a function of both height and weight and wherein the fit consists of a 3rd order function of weight combined with a linear function of height. The addition of height in this case did *not* improve the fit for small subjects.
(PDF)

S4 Fig. Good fit of liver training data, log inverse concentration as function of both height and weight. Three-dimensional scatter plot of data from the training set showing normal liver reciprocal mean concentration in units of ml/%ID shown on a log scale and plotted as a function of patient height and weight along with a surface showing the model A prediction. (PDF)

S5 Fig. Spleen test data, SUVs as a function of weight, height and age. For the independent test data, these scatter plots compare the correlations in normal spleen SUV_{bw} (column A, E, I), SUV_{lbm} (column B, F, J), SUV_{bsa} (column C, G, K) and SUV_{fdg} (column D, H, L) measurements with weight (row A, B, C, D), height (row E, F, G, H) and age (row I, J, K, L). Note, spleen concentrations were not measured in the training cohort and played no part in determining the BHN function used to calculate these SUV_{fdg} values. (PDF)

S6 Fig. Brain test data, SUVs as a function of weight, height and age. For the brain-only independent test data, these scatter plots compare the correlations in normal frontal gray matter SUV_{bw} (column A, E, I), SUV_{lbm} (column B, F, J), SUV_{bsa} (column C, G, K) and SUV_{fdg} (column D, H, L) measurements with weight (row A, B, C, D), height (row E, F, G, H) and age (row I, J, K, L). Note, brain concentrations were not measured in the training cohort and played no part in determining the BHN function used to calculate these SUV_{fdg} values. The lines and associated parameters seen in the legends were fitted to data from only the adult (>18 y) patients. The SUV_{fdg} values suggest a significant difference in brain glucose metabolism between adult and pediatric populations. In all graphs, triangles depict male patients, x's refer to female patients and o's are children under the age of 18. (PDF)

Acknowledgments

We wish to gratefully acknowledge the kind assistance and many productive conversations regarding the statistics related aspects of this work provided by Zhigang Zhang of the Memorial Sloan Kettering Cancer Center's department of Biostatistics.

Author Contributions

Conceptualization: Bradley J. Beattie, Tim J. Akhurst.

Data curation: Bradley J. Beattie, Finn Augensen.

Formal analysis: Bradley J. Beattie.

Funding acquisition: John L. Humm.

Investigation: Bradley J. Beattie, Finn Augensen.

Methodology: Bradley J. Beattie, Tim J. Akhurst.

Project administration: Bradley J. Beattie, John L. Humm.

Resources: John L. Humm.

Software: Bradley J. Beattie.

Supervision: Bradley J. Beattie, Tim J. Akhurst.

Validation: Bradley J. Beattie.

Visualization: Bradley J. Beattie.

Writing – original draft: Bradley J. Beattie.

Writing – review & editing: Bradley J. Beattie, Tim J. Akhurst.

References

1. Beattie BJ. Proposed changes to the ACR phantom filling procedure for more accurate and consistent activity concentrations. *J Appl Clin Med Phys*. 2019; 20(2):154–6. WOS:000458309000018. <https://doi.org/10.1002/acm2.12531> PMID: 30652408
2. Keyes JW Jr. SUV: standard uptake or silly useless value? *J Nucl Med*. 1995; 36(10):1836–9. PMID: 7562051.
3. Kim CK, Gupta NC, Chandramouli B, Alavi A. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med*. 1994; 35(1):164–7. PMID: 8271040.
4. Kim CK, Gupta NC. Dependency of standardized uptake values of fluorine-18 fluorodeoxyglucose on body size: comparison of body surface area correction and lean body mass correction. *Nucl Med Commun*. 1996; 17(10):890–4. <https://doi.org/10.1097/00006231-199610000-00011> PMID: 8951911.
5. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009; 50 Suppl 1:122S–50S. <https://doi.org/10.2967/jnumed.108.057307> PMID: 19403881; PubMed Central PMCID: PMC2755245.
6. Sarikaya I, Albatineh AN, Sarikaya A. Revisiting Weight-Normalized SUV and Lean-Body-Mass-Normalized SUV in PET Studies. *Journal of nuclear medicine technology*. 2020; 48(2):163–7. Epub 2019/10/13. <https://doi.org/10.2967/jnmt.119.233353> PMID: 31604893.
7. Erselcan T, Turgut B, Dogan D, Ozdemir S. Lean body mass-based standardized uptake value, derived from a predictive equation, might be misleading in PET studies. *Eur J Nucl Med Mol Imaging*. 2002; 29(12):1630–8. <https://doi.org/10.1007/s00259-002-0974-3> PMID: 12458398.
8. Halsne T, Muller EG, Spiten AE, Sherwani AG, Gyland Mikalsen LT, Revheim ME, et al. The Effect of New Formulas for Lean Body Mass on Lean-Body-Mass-Normalized SUV in Oncologic (18)F-FDG PET/CT. *Journal of nuclear medicine technology*. 2018; 46(3):253–9. <https://doi.org/10.2967/jnmt.117.204586> PMID: 29599401.
9. Kim WH, Kim CG, Kim DW. Comparison of SUVs Normalized by Lean Body Mass Determined by CT with Those Normalized by Lean Body Mass Estimated by Predictive Equations in Normal Tissues. *Nuclear medicine and molecular imaging*. 2012; 46(3):182–8. <https://doi.org/10.1007/s13139-012-0146-8> PMID: 24900058; PubMed Central PMCID: PMC4043039.
10. Kim CG, Kim WH, Kim MH, Kim DW. Direct Determination of Lean Body Mass by CT in F-18 FDG PET/CT Studies: Comparison with Estimates Using Predictive Equations. *Nuclear medicine and molecular imaging*. 2013; 47(2):98–103. <https://doi.org/10.1007/s13139-013-0207-7> PMID: 24900089; PubMed Central PMCID: PMC4041979.
11. Devriese J, Beels L, Maes A, Van De Wiele C, Gheysens O, Pottel H. Review of clinically accessible methods to determine lean body mass for normalization of standardized uptake values. *The quarterly journal of nuclear medicine and molecular imaging: official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the So*. 2016; 60(1):1–11. PMID: 26576735.
12. Villa C, Primeau C, Hesse U, Hougen HP, Lynnerup N, Hesse B. Body surface area determined by whole-body CT scanning: need for new formulae? *Clin Physiol Funct Imaging*. 2017; 37(2):183–93. <https://doi.org/10.1111/cpf.12284> PMID: 26302984.
13. Eskian M, Alavi A, Khorasanizadeh M, Viglianti BL, Jacobsson H, Barwick TD, et al. Effect of blood glucose level on standardized uptake value (SUV) in (18)F-FDG PET-scan: a systematic review and meta-analysis of 20,807 individual SUV measurements. *Eur J Nucl Med Mol Imaging*. 2019; 46(1):224–37. Epub 2018/10/24. <https://doi.org/10.1007/s00259-018-4194-x> PMID: 30350009.
14. Jahromi AH, Moradi F, Hoh CK. Glucose-corrected standardized uptake value (SUV(gluc)) is the most accurate SUV parameter for evaluation of pulmonary nodules. *Am J Nucl Med Mol Imaging*. 2019; 9(5):243–7. Epub 2019/11/28. PMID: 31772822; PubMed Central PMCID: PMC6872475.
15. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging*. 2015; 42(2):328–54. <https://doi.org/10.1007/s00259-014-2961-x> PMID: 25452219; PubMed Central PMCID: PMC4315529.
16. Quantitative Imaging Biomarkers Alliance. Quantitative FDG-PET Technical Committee. UPICT oncology FDG-PET CT protocol. [cited 2020 7/22/2020]. Available from: http://qibawiki.rsna.org/index.php?title=FDG-PET_tech_ctte.

17. Du Bois D, Du Bois EF. A formula to estimate the approximate surface area if height and weight be known. *Arch Intern Med.* 1916; 16(6):863–71. PubMed Central PMCID: PMC2520314.
18. O JH, Lodge MA, Wahl RL. Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0. *Radiology.* 2016; 280(2):576–84. <https://doi.org/10.1148/radiol.2016142043> PMID: 26909647; PubMed Central PMCID: PMC4976461.
19. Schwarz G. Estimating Dimension of a Model. *Ann Stat.* 1978; 6(2):461–4. WOS:A1978EQ63300014.
20. Akaike H. Fitting Autoregressive Models for Prediction. *Ann I Stat Math.* 1969; 21(2):243–&. WOS: A1969E435800002.
21. Theil H. Econometric Research in the Early 1950s - a Citation Classic Commentary on Economic Forecasts and Policy, by Theil,H. *Cc/Soc Behav Sci.* 1990;(17):24–. WOS:A1990CY30200001.
22. Bonferroni CE. *Teoria statistica delle classi e calcolo delle probabilità: Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*; 1936.
23. Nugent S, Tremblay S, Chen KW, Ayutyanont N, Roontiva A, Castellano CA, et al. Brain glucose and acetoacetate metabolism: a comparison of young and older adults. *Neurobiol Aging.* 2014; 35(6):1386–95. WOS:000333970800019. <https://doi.org/10.1016/j.neurobiolaging.2013.11.027> PMID: 24388785
24. Keramida G, Peters AM. FDG PET/CT of the non-malignant liver in an increasingly obese world population. *Clin Physiol Funct Imaging.* 2020. <https://doi.org/10.1111/cpf.12651> PMID: 32529712.
25. Zasadny KR, Wahl RL. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology.* 1993; 189(3):847–50. <https://doi.org/10.1148/radiology.189.3.8234714> PMID: 8234714.