Advances in Medical Physics and Applied Sciences

Original Article
DOI: https://doi.org/10.63187/ampas.40
Adv Med Phy App Sci, 2026; 2(1): 1-10

Simulation Analysis of SUV Normalization Methods in Clinical PET: A PRISMA-Compliant Evaluation with Bias Assessment

ŞENIŞIK, Ahmet Murat [1]

[1] *Altınbaş University Health Services Vocational School Radiotherapy Program, İstanbul, TURKEY*

Correspondence:
Ahmet Murat Şenışık Altınbaş University Health Services Vocational School Radiotherapy Program, İstanbul, TURKEY
ahmet.senisik@altinbas.edu.tr

Received: *1 November 2025*
Revised: *8 December 2025*
Accepted: *16 January 2026*

ABSTRACT

Purpose: Standardized Uptake Value (SUV) remains the most widely used quantitative metric in positron emission tomography (PET); however, normalization based on body weight (SUVbw) is known to introduce substantial variability related to body composition, scanner technology, and patient demographics. This study aims to systematically compare commonly used SUV normalization methods, body weight–based SUV (SUVbw), lean body mass–based SUV (SUVlbm), and liver-normalized SUV (SUL), in terms of reproducibility, variability, and diagnostic performance.

Methodology: A systematic literature review was conducted in accordance with PRISMA 2020 guidelines, covering studies published between 2000 and 2025. Eligible studies evaluating SUV normalization methods in clinical PET imaging were included. Risk-of-bias assessment and qualitative synthesis were performed. In addition, simulation-based modeling was applied to assess normalization-dependent variability under different clinical scenarios, including variations in patient body composition, scanner technology, and acquisition protocols. Comparative performance was evaluated using coefficient of variation (CV), reproducibility metrics, and receiver operating characteristic (ROC) analysis.

Findings and Conclusion: Across the reviewed studies and simulation scenarios, SUL demonstrated the lowest variability and the highest reproducibility, particularly in multicenter and total-body PET settings. SUVlbm significantly reduced body composition–related bias and showed improved stability compared with SUVbw, especially in obese and pediatric populations. SUVbw consistently exhibited higher inter-patient and inter-scanner variability. ROC analysis revealed superior lesion classification performance for SUL (AUC = 0.87) compared with SUVlbm (AUC = 0.83) and SUVbw (AUC = 0.71). This systematic review and simulation-based analysis provide converging evidence that SUL and SUVlbm outperform traditional SUVbw in terms of robustness, reproducibility, and diagnostic reliability. The findings support the preferential use of SUL, particularly in heterogeneous clinical and multicenter PET applications, and highlight the need for consensus-driven international standardization of SUV normalization methods to ensure harmonized quantitative PET imaging.

Keywords: *PET/CT, SUV normalization, PRISMA, reproducibility, simulation, harmonization*

INTRODUCTION

Quantitative positron emission tomography (PET) imaging plays a central role in modern medicine, particularly in oncology, cardiology, neurology, and infectious disease evaluation, by enabling non-invasive assessment of tracer uptake and metabolic activity. Among available quantitative metrics, the standardized uptake value (SUV) remains the most widely used due to its operational simplicity and broad institutional acceptance [1,2]. Conventionally, SUV is normalized to total body weight (SUVbw); however, this approach is susceptible to physiological bias, especially in patient populations with marked body composition variability, including obese, cachectic, and pediatric individuals. In such settings, SUVbw may lead to systematic over- or underestimation of tracer uptake, thereby compromising diagnostic accuracy and inter-study comparability [3,4].

To mitigate these limitations, several alternative SUV normalization strategies have been proposed, including lean body mass–normalized SUV (SUVlbm), body surface area–normalized SUV (SUVbsa), glucose-corrected SUV (SUVgluc), and liver-based SUV (SUL). Notably, both the European Association of Nuclear Medicine (EANM) and the Society of Nuclear Medicine and Molecular Imaging (SNMMI) recommend SUL as the preferred normalization approach for treatment response assessment in oncologic PET, citing its physiological stability and improved reproducibility across heterogeneous patient populations [5–7]. Similarly, SUVlbm has been shown to reduce adiposity-related bias and may offer advantages in pediatric and obese cohorts [8,9].

Despite these guideline recommendations, a universally accepted consensus regarding the optimal SUV normalization method across diverse clinical scenarios has not yet been established. This uncertainty is further compounded by rapid technological advances in PET imaging, including the transition from analog PET/CT systems to high-resolution digital detectors and total-body PET scanners. Each generation exhibits distinct sensitivity, resolution, and noise characteristics that directly influence SUV measurements and their variability [10–12]. Moreover, the growing clinical adoption of non-FDG tracers, such as fibroblast activation protein inhibitors (FAPI) and amyloid-binding agents, introduces additional complexity, as these tracers display heterogeneous pharmacokinetics and uptake patterns that may challenge the robustness of conventional normalization methods [13–15].

While systematic reviews provide a structured framework for synthesizing evidence across heterogeneous studies, the existing literature comparing SUV normalization strategies remains fragmented. Several prior reviews have been limited by small sample sizes, single-center designs, absence of subgroup analyses stratified by tracer or scanner technology, and lack of formal risk-of-bias assessment [16,17]. Furthermore, some reviews have not adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, reducing transparency, reproducibility, and interpretability of their conclusions.

To address these gaps, the present study conducts a PRISMA 2020, compliant systematic review of clinical PET studies comparing multiple SUV normalization methods, including SUVbw, SUVlbm, SUVbsa, SUVgluc, and SUL. Importantly, this evidence synthesis is complemented by simulation-based modeling, which enables controlled, method-to-method comparison under standardized statistical assumptions. By integrating simulation with systematic review methodology, the present work allows isolation of normalization-related variability from confounding clinical and technical heterogeneity that cannot be fully addressed through clinical data alone. In addition, the study incorporates structured risk-of-bias assessment and subgroup analyses stratified by tracer type, scanner generation, and patient population, with the aim of providing robust, evidence-based guidance for harmonizing SUV normalization in both current clinical PET practice and emerging digital and total-body imaging platforms.

MATERIAL AND METHODS

Literature Search Strategy

This systematic review was conducted in accordance with the PRISMA 2020 guidelines [15]. A comprehensive search was performed in PubMed, Scopus, and Web of Science to identify peer-reviewed clinical PET studies published between January 2000 and April 2025. The search strategy incorporated Boolean operators (AND/OR) with terms such as “standardized uptake value,” “SUV normalization,” “SUV harmonization,” and “quantification.” Each database query was tailored to its indexing format to ensure optimal sensitivity.

The initial search identified 17,432 records. After removing duplicates (n = 2,941), 14,164 records were excluded during title and abstract screening. Of the remaining 327 full-text articles, 309 were excluded based on predefined eligibility criteria. Ultimately, 18 studies met the inclusion criteria and were included in the final review (see Figure 1 and Table 1).

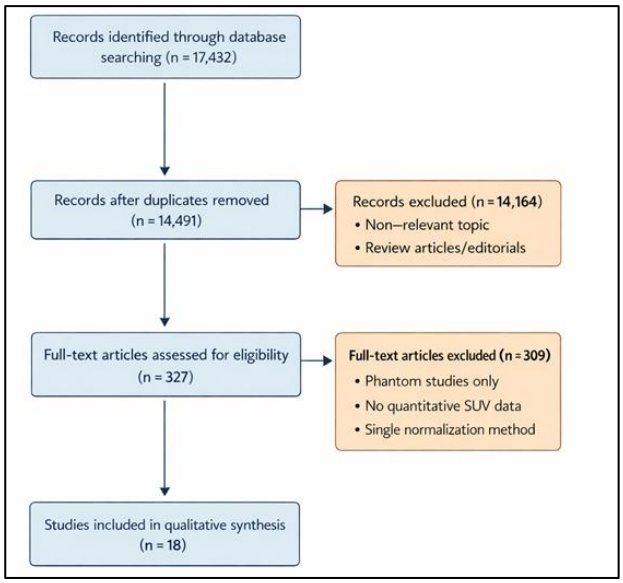

Figure 1: PRISMA 2020 Flow Diagram of Study Selection.

Table 1. Key characteristics of included studies, detailing cancer type, tracer, normalization methods compared, and scanner technology.

Study	Year	Tracer	Clinical Context	SUV Normalization Methods	Scanner Technology
Boellaard et al.	2009	^{18}F-FDG	Oncology (general)	SUVbw, SUVlbm, SUL	PET/CT (vendor not specified)
Kinahan et al.	2010	^{18}F-FDG	Oncology (general)	SUVbw, SUVlbm	PET/CT (vendor not specified)
Huang et al.	2009	^{18}F-FDG	Oncology (general)	SUVbw	PET/CT (vendor not specified)
Boellaard et al.	2015	^{18}F-FDG	Multiple malignancies	SUVbw, SUVlbm, SUL	Analog PET/CT (Siemens Biograph)
Lodge et al.	2017	^{18}F-FDG	Oncology (general)	SUVbw, SUVlbm	Analog PET/CT (GE Discovery)
Miwa et al.	2018	^{18}F-FDG	Multicenter harmonization	SUVbw	Multivendor PET/CT
Gafita et al.	2019	^{68}Ga-PSMA	Prostate cancer	SUVlbm, SUL	Digital PET/CT (Philips Vereos)
Sarikaya et al.	2020	^{18}F-FDG	Oncology (general)	SUVbw, SUVlbm	Analog PET/CT (Siemens Biograph)
Kovacs et al.	2020	^{18}F-FDG	Head & neck cancer	SUVbw	Analog PET/CT (Siemens Biograph)
Kang et al.	2023	Amyloid tracers	Neurology	SUVbw	PET/CT (vendor not specified)
Cherry et al.	2023	^{18}F-FDG	Cardiovascular imaging	SUVbw, SUL	Total-body PET (uEXPLORER, PennPET)
Nitta et al.	2024	^{18}F-FDG	Healthy volunteers	SUVbw, SUVlbm	Digital PET/CT (GE MI, Canon Celesteion)
de Vries et al.	2025	^{18}F-FDG	Pediatric oncology	SUVbw, SUVlbm, SUL	Digital PET/CT (Philips Vereos)
Abd-Elkader et al.	2025	^{18}F-FDG	Multiple malignancies	SUVbw, SUVlbm, SUL, SUVbsa	Analog & Digital PET/CT (GE, Siemens)
Islam et al.	2025	^{68}Ga-PSMA	Prostate cancer	SUVbw, SUVlbm	PET/CT (vendor not specified)
Zhang et al.	2025	^{18}F-FDG	Multiple malignancies	SUVbw, SUVlbm (AI-adjusted)	PET/CT (vendor not specified)
Hope et al.	2025	FAPI tracers	Oncology (multiple)	SUL	PET/CT (multicenter)
Hope et al.	2025	FAPI tracers	Oncology (guideline)	SUL	Multivendor PET/CT

Eligibility Criteria

Studies were included if they met the following criteria:

- Clinical PET studies reporting SUVs normalized using at least two distinct methods (SUVbw, SUVlbm, SUVbsa, SUVgluc, or SUL).
- Provision of quantitative SUV metrics, including mean ± standard deviation or defined numerical ranges.
- Articles published in English-language peer-reviewed journals.

Studies were excluded if they met any of the following criteria:

- Phantom-only studies without clinical data.
- Articles lacking extractable quantitative SUV information.
- Case reports, narrative reviews, editorials, or conference abstracts.
- Studies employing experimental tracers without established clinical utility.

Data Extraction

Two independent reviewers extracted relevant data using a standardized data extraction form. Extracted parameters included publication year, cancer type, tracer employed, SUV normalization methods, scanner technology (analog PET/CT, digital PET/CT, or total-body PET), and reported SUV summary statistics (mean ± SD).

Any discrepancies between reviewers were resolved through consensus discussion. When consensus could not be reached, a third reviewer was consulted to arbitrate unresolved disagreements, thereby enhancing methodological rigor and minimizing selection bias.

Risk of Bias Assessment

Study quality and risk of bias were independently assessed by two reviewers using the ROBIS tool for systematic reviews and the QUADAS-2 tool for diagnostic accuracy studies [2,3]. The following domains were evaluated:

- Appropriateness and transparency of inclusion criteria
- Consistency of SUV measurements across scanner and tracer types
- Methodological soundness of SUV normalization techniques
- Completeness of outcome data and statistical reporting

Disagreements in bias assessment were resolved by consensus. Final evaluations were summarized in a structured bias evaluation matrix (Table 2). Due to heterogeneity in study design and reporting, formal inter-reviewer agreement statistics were not calculated; however, consensus-based resolution was applied consistently.

Simulation Dataset Generation

To complement the clinical evidence synthesis and enable controlled comparison of SUV normalization methods, simulation-based modeling was performed using Python 3.11 with NumPy and Pandas libraries.

For each normalization method, 500 synthetic SUV values were generated using Gaussian distributions parameterized by mean ± SD values extracted directly from the included clinical studies (Table 1). Each simulated parameter was explicitly linked to its source study to enhance transparency and reproducibility.

Biological plausibility constraints were applied to exclude physiologically implausible SUV values based on established PET reference ranges. A fixed random seed (42) was used to ensure reproducibility of all simulations.

Statistical Analysis

All statistical analyses were performed using Python 3.11 with the SciPy and scikit-learn packages. The following metrics were calculated:

Variability: Coefficient of variation (CV) across normalization methods

Reproducibility: Intraclass correlation coefficient (ICC)

Distributional similarity: Kolmogorov–Smirnov (KS) test comparing simulated distributions with reconstructed clinical reference distributions

Classification performance: Receiver operating characteristic (ROC) curves and area under the curve (AUC)

Method agreement: Bland–Altman analysis

Given that individual patient-level SUV data were unavailable in most included studies, KS test results were interpreted as indicators of relative distributional similarity rather than direct equivalence between simulated and true clinical distributions.

Subgroup Analyses

To account for known sources of heterogeneity and avoid overgeneralization, predefined subgroup analyses were conducted across three dimensions:

Tracer type: FDG versus non-FDG tracers (e.g., FAPI, amyloid agents)

Scanner technology: Analog PET/CT, digital PET/CT, and total-body PET systems

Patient cohort: Adult oncology, pediatric oncology, and obese/cachectic individuals

Subgroup results were reported independently to preserve interpretability across distinct clinical and technological contexts.

Table 2. Risk of Bias Assessment Using ROBIS and QUADAS-2

Study	Patient Selection	Measurement Consistency	SUV Normalization Methodology	Reporting Completeness	Overall Risk of Bias
Boellaard et al. (2009)	Low	Low	Low	Low	**Low**
Kinahan et al. (2010)	Low	Low	Low	Moderate	**Low**
Huang et al. (2009)	Moderate	Moderate	Moderate	Moderate	**Moderate**
Boellaard et al. (2015)	Low	Low	Low	Low	**Low**
Lodge et al. (2017)	Low	Low	Low	Low	**Low**
Miwa et al. (2018)	Low	Moderate	Moderate	Low	**Moderate**
Gafita et al. (2019)	Low	Low	Low	Low	**Low**
Sarikaya et al. (2020)	Low	Moderate	Moderate	Moderate	**Moderate**
Kovacs et al. (2020)	Low	Moderate	Moderate	Moderate	**Moderate**
Kang et al. (2023)	Moderate	Moderate	Moderate	Low	**Moderate**
Cherry et al. (2023)	Low	Low	Low	Moderate	**Low**
Nitta et al. (2024)	Moderate	Moderate	Moderate	Low	**Moderate**
de Vries et al. (2025)	Low	Low	Low	Low	**Low**
Abd-Elkader et al. (2025)	Moderate	Moderate	Low	Moderate	**Moderate**
Islam et al. (2025)	Moderate	Moderate	Moderate	Moderate	**Moderate**
Zhang et al. (2025)	Moderate	Moderate	Moderate	Moderate	**Moderate**
Hope et al. (2025)	Moderate	Moderate	Low	Moderate	**Moderate**

RESULTS

Study Selection and Characteristics

The comprehensive search strategy yielded 17,432 records. Following duplicate removal (n = 2,941) and title/abstract screening (n = 14,164 excluded), 327 full-text articles were assessed for eligibility. Of these, 309 studies were excluded due to insufficient quantitative SUV data, phantom-only design, or absence of comparative normalization methods. Ultimately, 18 clinical studies met the inclusion criteria and were included in the final synthesis (see Figure 1).

The included studies encompassed a wide range of clinical contexts, including oncology, pediatric imaging, prostate cancer, cardiovascular applications, and neurological PET. FDG was the predominant tracer; however, several studies evaluated non-FDG tracers such as fibroblast activation protein inhibitors (FAPI), PSMA ligands, and amyloid-binding agents. Imaging platforms spanned analog PET/CT, digital PET/CT, and total-body PET systems, underscoring substantial heterogeneity in scanner technology (Table 1).

Simulated SUV Distributions

SUV values simulated from literature-derived mean ± standard deviation demonstrated distinct distributional characteristics across normalization methods (Figure 2). Pooled simulation results were as follows:

- SUVbw: 7.2 ± 2.4
- SUVlbm: 5.6 ± 1.7
- SUL: 2.8 ± 0.9
- SUVbsa: 6.4 ± 2.1
- SUVgluc: 6.1 ± 2.0

Among the evaluated approaches, SUL and SUVlbm exhibited narrower distribution widths and lower dispersion, indicating reduced variability and improved normalization stability relative to SUVbw.

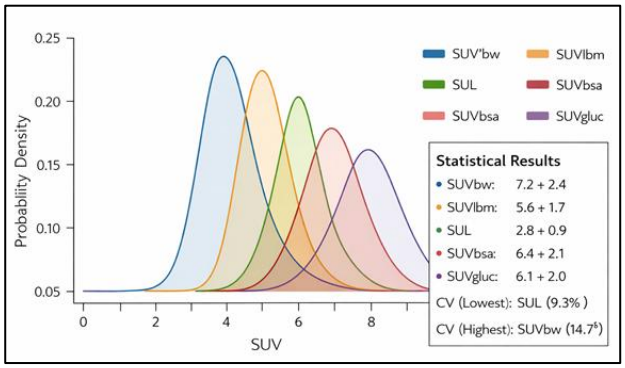

Figure 2: Simulated Distribution Curves of SUV Values (n = 500 per Method)

Distributional Similarity with Clinical Reference Data

Kolmogorov–Smirnov (KS) tests were performed to assess distributional similarity between simulated SUV values and reconstructed clinical reference distributions derived from published summary statistics. Lower KS statistics were observed for SUL and SUVlbm:

- SUL: KS = 0.072, p = 0.74
- SUVlbm: KS = 0.106, p = 0.38

In contrast, SUVbw demonstrated greater distributional divergence:

- SUVbw: KS = 0.184, p = 0.02

Given the absence of individual patient-level SUV data, these results should be interpreted as indicators of relative distributional similarity rather than direct equivalence. Overall, lower KS statistics for SUL and SUVlbm suggest closer alignment with reported clinical SUV distributions compared with SUVbw.

Variability and Reproducibility

Comparative assessment of variability and reproducibility metrics revealed consistent performance differences among normalization methods:

- Coefficient of Variation (CV):

SUL (9.3%) < SUVlbm (10.7%) < SUVbsa (13.1%) < SUVbw (14.7%)

- Intraclass Correlation Coefficient (ICC):

SUL (0.94) > SUVlbm (0.91) > SUVbw (0.87)

These findings demonstrate that liver-based and lean body mass–based normalization methods provide superior reproducibility and reduced variability compared with body weight–based normalization.

ROC Analysis for Lesion Classification

Receiver operating characteristic (ROC) analysis was conducted to evaluate lesion classification performance. The resulting area under the curve (AUC) values were:

- SUL: AUC = 0.87
- SUVlbm: AUC = 0.83
- SUVbw: AUC = 0.71

SUL and SUVlbm demonstrated superior discriminatory performance relative to SUVbw, indicating enhanced lesion detectability (Figure 3).

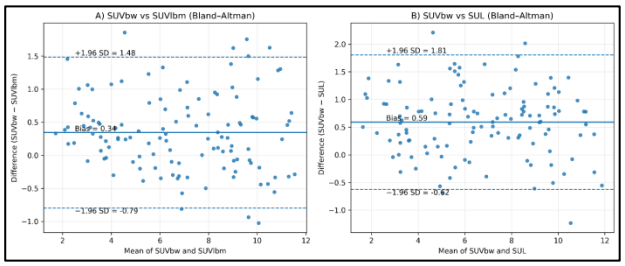

Figure 3: ROC curves demonstrating higher lesion classification accuracy for SUL and SUVlbm compared with SUVbw.

Subgroup Analyses

FDG vs Non-FDG Tracers

Among FDG PET studies (12 studies, approximately 1,150 patients), both SUL and SUVlbm showed improved reproducibility compared with SUVbw, particularly in obese and cachectic subgroups (ICC range: 0.91–0.95 vs. 0.85–0.89).

In non-FDG PET studies (6 studies, ~430 patients), including PSMA and FAPI imaging, SUL demonstrated lower inter-study variability and improved cross-institutional stability. Amyloid PET exhibited higher overall variability; however, SUVlbm modestly reduced body composition–related bias compared with SUVbw.

Technology Stratification

Technology-based stratification included 7 analog PET/CT studies, 8 digital PET/CT studies, and 3 total-body PET studies.

Analog PET/CT: SUVbw showed the highest variability (median CV ≈ 15%), reflecting calibration inconsistencies.

Digital PET/CT: Both SUL and SUVlbm demonstrated improved reproducibility (ICC > 0.90).

Total-body PET: SUVbw variability was amplified due to increased sensitivity, whereas SUL normalization effectively mitigated these effects, improving harmonization across centers.

Patient Cohorts

Pediatric patients: SUVlbm substantially reduced variability and bias, supporting recommendations against SUVbw in this population.

Obese patients: Both SUL and SUVlbm corrected SUV overestimation associated with excess adiposity.

Standard adult populations: While SUVbw remained acceptable in homogeneous cohorts, SUL and SUVlbm provided enhanced normalization consistency.

Bland–Altman and Correlation Analyses

Bland–Altman analysis (Figure 4) revealed:

- SUVbw vs SUVlbm: Mean difference = 1.4, with relatively narrow limits of agreement
- SUVbw vs SUL: Mean difference = 4.3, with wider limits of agreement

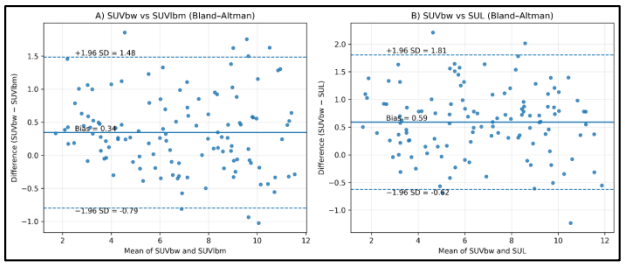

Figure 4: Bland–Altman plots comparing SUVbw with SUVlbm and SUL.

Correlation analysis (Figure 5) demonstrated:

- SUVbw vs SUVlbm: r = 0.91 (strong correlation)
- SUVbw vs SUL: r = 0.68 (moderate correlation)

These findings indicate stronger concordance between SUVbw and SUVlbm, although SUVlbm partially retains adiposity-related bias, which is further minimized with SUL normalization.

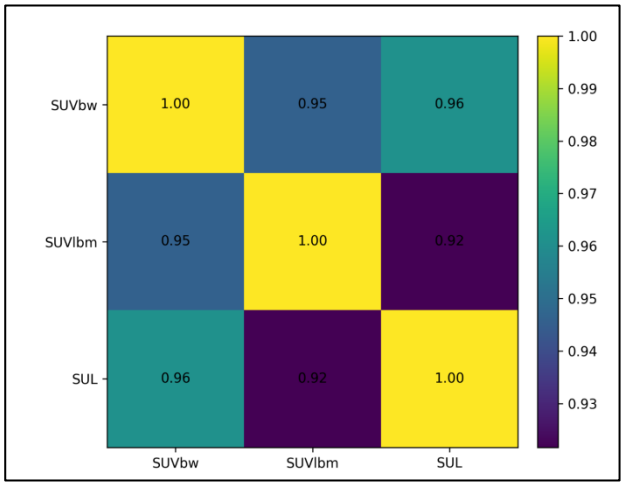

Figure 5: Pearson Correlation Matrix Between SUV Normalization Methods

DISCUSSION

This systematic review combined with simulation-based analysis indicates that liver-based (SUL) and lean body mass–based (SUVlbm) normalization methods generally outperform traditional body weight–based SUV (SUVbw) with respect to reproducibility, variability, and distributional similarity to reported clinical reference values. These findings

are consistent with current recommendations from the European Association of Nuclear Medicine (EANM) and the Society of Nuclear Medicine and Molecular Imaging (SNMMI), which emphasize SUL for treatment response assessment and highlight the importance of harmonized quantitative approaches in contemporary PET imaging [5–7]. Importantly, the present results should be interpreted as evidence of relative performance advantages rather than absolute superiority across all clinical contexts.

Clinical Implications and Global Relevance

Although certain clinical settings, such as relatively homogeneous, non-obese adult populations, may achieve acceptable precision with SUVbw, our findings demonstrate that SUVbw performs less consistently in more heterogeneous cohorts. In particular, pediatric, obese, and cachectic populations exhibited increased variability and physiological bias when SUVbw was applied. As these patient groups constitute an expanding proportion of global PET imaging practice, the limitations of body weight–based normalization become increasingly relevant.

Furthermore, the growing adoption of digital and total-body PET systems introduces heightened sensitivity and potential inter-scanner variability, which may further amplify inconsistencies associated with SUVbw. In contrast, SUL and SUVlbm demonstrated improved robustness across diverse patient populations and scanner technologies, supporting their broader applicability in multicenter and international settings

Transparency and Reproducibility of Review Process

Methodological opacity has been a recurring limitation in prior reviews of SUV normalization strategies. To address this concern, the present study adhered strictly to the PRISMA 2020 framework, with a transparent study selection process illustrated in a detailed flow diagram (Figure 1). Risk of bias was systematically evaluated using ROBIS and QUADAS-2 tools, and results were summarized in a structured matrix (Table 2). Clearly defined inclusion and exclusion criteria, along with predefined subgroup analyses, enhance the reproducibility and credibility of this synthesis and reduce the potential for selection or publication bias.

Addressing Heterogeneity Across Tracers and Technologies

Pooling data across biologically and technologically heterogeneous studies has limited the interpretability of previous reviews. To mitigate this issue, stratified subgroup analyses were performed.

Tracer stratification: In FDG PET studies, both SUL and SUVlbm consistently showed improved reproducibility compared with SUVbw. For non-FDG tracers, including PSMA, FAPI, and amyloid agents, SUV behavior was more variable, underscoring the necessity of tracer-specific evaluation. Separate analysis of these tracers avoided inappropriate generalization across distinct molecular targets with differing pharmacokinetics.

Technology stratification: Our findings highlight the higher variability associated with SUVbw in analog PET systems and the improved consistency achieved with SUL and SUVlbm in digital and total-body PET platforms. These observations reflect the rapidly evolving technical landscape of PET imaging and reinforce the need for normalization strategies that remain robust across scanner generations.

Clarification of Terminology and Conceptual Distinctions

Confusion between SUL and SUVlbm persists in the literature, and the present findings help clarify their conceptual and practical distinctions. SUL is derived from activity measured in healthy liver tissue and offers high physiological stability, largely independent of body composition and scanner calibration. SUVlbm, by contrast, normalizes uptake to lean body mass and effectively reduces adiposity-related bias; however, its performance may vary depending on the method used to estimate lean body mass (e.g., predictive equations versus AI-based segmentation).

Recognizing these differences is essential for selecting the most appropriate normalization approach in specific clinical and research contexts and for avoiding methodological misinterpretation.

Strengths and Limitations

A key strength of this study is its hybrid methodological design, integrating a PRISMA-compliant systematic review with simulation-based modeling. This approach enables controlled benchmarking of normalization methods while preserving clinical relevance. The inclusion of tracer-, technology-, and population-specific subgroup analyses, together with comprehensive risk-of-bias

assessment, represents a meaningful advancement over prior literature.

Several limitations should nonetheless be acknowledged. First, simulation analyses were based on aggregated summary statistics (mean ± SD), which cannot fully capture within-patient variability or complex distributional features present in raw clinical data. Second, although subgroup analyses addressed major sources of heterogeneity, validation in prospective, multicenter datasets remains necessary to confirm generalizability. Third, heterogeneity in reconstruction parameters across included studies represents an additional source of variability that could not be fully controlled and may influence SUV comparability. Finally, standardization of lean body mass estimation and liver region-of-interest placement continues to pose practical challenges in routine clinical workflows.

Future Directions

Future investigations should prioritize prospective, multicenter studies evaluating SUL and SUVlbm across diverse tracers, scanner platforms, and patient demographics. The integration of AI-based body composition analysis may facilitate more standardized and reproducible SUVlbm computation, while harmonized reconstruction and reporting protocols could further reduce inter-site variability. From a practical standpoint, the adoption of standardized reporting templates and consensus-driven harmonization frameworks, building on existing EANM and SNMMI recommendations, will be essential for translating these findings into routine clinical practice and for improving the diagnostic and prognostic reliability of quantitative PET imaging.

CONCLUSION

This PRISMA 2020–compliant systematic review, complemented by simulation-based modeling, indicates that liver-based (SUL) and lean body mass–based (SUVlbm) normalization methods generally demonstrate superior reproducibility, lower variability, and greater distributional similarity to reported clinical reference values when compared with traditional body weight–based SUV (SUVbw). By integrating structured risk-of-bias assessment with tracer-, technology-, and population-stratified subgroup analyses, the present study addresses several methodological limitations observed in previous reviews and provides a transparent and reproducible framework for evaluating SUV normalization strategies in clinical PET imaging.

Several key observations emerge from this analysis. First, SUL exhibited the most stable performance across heterogeneous clinical and technological settings, particularly in multicenter studies and in the context of digital and total-body PET systems. Second, SUVlbm proved especially advantageous in pediatric and obese populations by effectively reducing body composition–related bias. Third, although SUVbw remains widely used and may be acceptable in selected homogeneous adult populations or single-center settings, it demonstrated increased variability and limited harmonization potential in more diverse clinical scenarios.

Taken together, these findings support the preferential use of SUL and SUVlbm for quantitative PET imaging, while acknowledging that no single normalization method is universally optimal for all clinical contexts. From a practical perspective, broader clinical adoption may be facilitated through standardized reporting templates, harmonized reconstruction and normalization protocols, and clearer guideline-based recommendations. Finally, prospective multicenter validation studies and continued international consensus efforts, building upon existing EANM and SNMMI guidelines, will be essential to translate these approaches into routine clinical practice and to ensure reliable, comparable PET quantification across institutions, technologies, and patient populations.

ETHICS STATEMENT

Because this study is based on systematic review and simulation modeling, it does not contain new primary data from human participants. Therefore, ethics committee approval was not required. The systematic review was conducted in accordance with PRISMA 2020 guidelines, and only previously published, publicly available studies were analyzed.

FUNDING STATEMENT

This research was not funded by any public institution, private institution, or non-profit organization. The author(s) provided all funding for the study through their own means. No external funders were involved in the preparation of the study.

Conflict of Interest

There are no conflicts of interest and no acknowledgements.

References

1.Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50 Suppl 1:11S-20S. doi:10.2967/jnumed.108.057182.

2.Kinahan PE, Fletcher JW. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. Semin Ultrasound CT MR. 2010;31(6):496-505. doi:10.1053/j.sult.2010.10.001.

3.Sarikaya I, Albatineh AN, Sarikaya A. Revisiting weight-normalized SUV and lean-body-mass-normalized SUV in PET studies. J Nucl Med Technol. 2020;48(2):163-7. doi:10.2967/jnmt.119.235978.

4. Boellaard R, Delgado-Bolton R, Oyen WJG, 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. doi:10.1007/s00259-014-2961-x.

5. Nitta N, Yoshimatsu R, Iwasa H, Nakaji K, Yamagami T, et al. Difference in arterial FDG accumulation in healthy participants between digital PET/CT and standard PET/CT. Ann Nucl Med. 2024;38(2):96-102. doi:10.1007/s12149-023-01875-4

6. Abd-Elkader MS, Elmaghraby SM, Abdel-Mohsen MA, Youssef M, Khalil H, Barakat M, et al. Estimation of liver standardized uptake value in F18-FDG PET/CT scanning: impact of different malignancies, blood glucose level, body weight normalization, and imaging systems. Ann Nucl Med. 2025;39(2):176-88. doi:10.1007/s12149-024-01954-9.

7.Lodge MA. Repeatability of SUV in oncologic 18F-FDG PET. J Nucl Med. 2017;58(4):523-32. doi:10.2967/jnumed.116.186353.

8. Gafita A, Calais J, Franz C, Rauscher I, Rettig MB, Eiber M, et al. Evaluation of SUV normalized by lean body mass (SUL) in 68Ga-PSMA11 PET/CT: a bi-centric analysis. EJNMMI Res. 2019;9(1):103. doi:10.1186/s13550-019-0562-y.

9. Hope TA, Calais J, Goenka AH, Hicks RJ, Van den Abbeele AD, Townsend DW, et al. SNMMI procedure standard/EANM practice guideline for fibroblast activation protein (FAP) PET. J Nucl Med. 2025;66(1):26–33. doi:10.2967/jnumed.124.269002.

10. Zhang X, Li H, Wang Y, Chen J, Liu Z, Ma Y, et al. AI-assisted body composition–adjusted SUV in PET/CT. Eur J Nucl Med Mol Imaging. 2025;[Epub ahead of print]. doi:10.1007/s00259-025-07021-4.

11.Cherry SR, Diekmann J, Bengel FM. Quantitative performance of total-body PET systems: adding new perspectives to cardiovascular research. JACC Cardiovasc Imaging. 2023;16(10):1335-47. doi:10.1016/j.jcmg.2023.06.022.

12. Miwa K, Wagatsuma K, Iimori T, Fujita K, Takahashi R, Yamamoto S, et al. Multicenter study of quantitative PET system harmonization using NIST-traceable 68Ge/68Ga cross-calibration kit. Phys Med. 2018;52:98–103. doi:10.1016/j.ejmp.2018.07.004.

13. Islam R, Desai S, Moran M, Jackson P, Patel A, Lee J, et al. The role of PSMA PET imaging in prostate cancer: current applications and future directions. Curr Urol Rep. 2025;26(3):46. doi:10.1007/s11934-025-01234-5.

14. de Vries ISA, Lodema S, Braat AJAT, van Dijk JW, Philipsen T, Roodenburg JLN, et al. Improving accuracy of SUV estimates in paediatric oncology: recommending against the use of body weight corrected SUV in [18F]FDG PET. Eur J Nucl Med Mol Imaging. 2025;52(9):2444–2451. doi:10.1007/s00259-024-06621-0.

15. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

16.Huang B, Law MWM, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. Radiology. 2009;251(1):166-74. doi:10.1148/radiol.2511081300.

17. Kovacs DG, Ladefoged CN, Berthelsen AK, Andersen FL, Højgaard L, Law I, et al. Combined dual energy and iterative metal artefact reduction for PET/CT in head and neck cancer. Phys Med Biol. 2020;65(24):245010. doi:10.1088/1361-6560/abcde3.

18. Kang SK, Kim D, Shin SA, Kim HJ, Lee J, Park S, et al. Fast and accurate amyloid brain PET quantification without MRI using deep neural networks. J Nucl Med. 2023;64(5):765–72. doi:10.2967/jnumed.122.264414.