1-12-2023

Evaluation of Automated Anthropometrics Produced By Smartphone-Based Machine Learning: A Comparison With Traditional Anthropometric Assessments

Austin J. Graybeal
University of Southern Mississippi, Austin.Graybeal@usm.edu

Caleb F. Brandner
University of Southern Mississippi

Grant M. Tinsley
Texas Tech University, grant.tinsley@ttu.edu

Evaluation of automated anthropometrics produced by smartphone-based machine learning: a comparison with traditional anthropometric assessments

Austin J. Graybeal[1], Caleb F. Brandner[1], Grant M. Tinsley[2]

[1]School of Kinesiology & Nutrition, College of Education and Human Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA. Austin.graybeal@usm.edu; Caleb.Brandner@usm.edu

[2]Department of Kinesiology & Sport Management, Texas Tech University, Lubbock, TX 79409, USA. grant.tinsley@ttu.edu

Corresponding Author: Austin J. Graybeal, PhD, CSCS, Assistant Professor, School of Kinesiology & Nutrition, College of Education and Human Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA. Phone: 601-266-5996. Email address: austin.graybeal@usm.edu. ORCID: 0000-0003-4520-9230

Short title: Smartphone-based automated anthropometrics

Accepted manuscript

ABSTRACT

Automated visual anthropometrics produced by mobile applications are accessible and cost-effective with the potential to assess clinically relevant anthropometrics without a trained technician present. Thus, the aim of this study was to evaluate the precision and agreement of smartphone-based automated anthropometrics against reference tape measurements. Waist and hip circumference (WC; HC), waist-to-hip ratio (WHR), and waist-to-height ratio (W:HT), were collected from 115 participants (69 F) using a tape measure and two smartphone applications (MeThreeSixty®, myBVI®) across multiple smartphone types. Precision metrics were used to assess test-retest precision of the automated measures. Agreement between the circumferences produced by each mobile application and the reference were assessed using equivalence testing and other validity metrics. All mobile applications across smartphone types produced reliable estimates for each variable with ICCs ≥0.93 (all p<0.001) and RMS-%CV between 0.5%-2.5%. PE for WC and HC were between 0.5cm-1.9cm. WC, HC, and W:HT estimates produced by each mobile application demonstrated equivalence with the reference tape measurements using 5% equivalence regions. Mean differences via paired t-tests were significant for all variables across each mobile application (all p<0.050) showing slight underestimation for WC and slight overestimation for HC which resulted in a lack of equivalence for WHR compared to the reference tape measure. Overall, the results of our study support the use of WC and HC estimates produced from automated mobile applications, but also demonstrates the importance of accurate automation for WC and HC estimates given their influence on other anthropometric assessments and clinical health markers.

Keywords: digital anthropometry, mobile health, body composition assessment, obesity

Accepted manuscript

INTRODUCTION

Obesity is rapidly increasing with disparately high rates for individuals who are of low socioeconomic status (SES) (1) or living in rural communities. (2,3) Specifically, both low SES (4,5) and rural occupancy (6) are associated with higher rates of abdominal obesity, which is itself linked to a higher risk of cardiometabolic abnormalities. (7) Traditionally, abdominal obesity is evaluated by standard anthropometric assessments that include circumferences of the waist -and -hips specifically, given their association with cardiometabolic health risks (8,9) and mortality. (10) Nevertheless, traditional anthropometric measures lack feasibility for those without access to clinical care, which is concerning given that there are few alternative methods that can successfully provide remote and cost-effective assessments without a trained technician present. Therefore, the development of remote healthcare tools that can provide accurate anthropometric assessments without additional costs are of critical importance.

Interestingly, the adaptations made to traditional healthcare models during the COVID-19 pandemic may have, unintentionally, provided a potential solution to this pressing issue. At the onset of the pandemic, healthcare facilities were forced to find remote alternatives to providing clinical care. Given that the majority of US adults own a smartphone (11) with a similar ownership existing across more vulnerable populations, (12,13) the demand for remote healthcare solutions led to the swift integration of mobile healthcare models that enable patients with access to care regardless of limited transportation or geographical location. In fact, recent evidence demonstrates the feasibility of mobile healthcare interventions in a rural setting showing reductions in waist circumference (WC) and visceral adiposity with this approach. (14) Additionally, and contrary to the observations made in a traditional healthcare setting, (15) intention to use and satisfaction with mobile health services is inversely associated with perceived health status which notes its utility for individuals of poor health. However, increased access to communication with healthcare providers is only one facet of the total healthcare experience. As such, the surge in mobile healthcare use coupled with the accelerated increase in obesity requires parallel advancements in remote health tools, such as automated anthropometrics, that can conveniently assess patient health status without additional costs.

Because virtually all smartphones can now employ high-resolution imaging through advances in smartphone camera technology, newly developed mobile applications can now

leverage machine learning to automate anthropometric assessments from as little as two self-taken images. Furthermore, this method has recently shown to be associated with traditional measurement methods, (16) agree with multi-compartment body composition estimation models, (17) and demonstrates higher predictive ability of visceral adipose tissue compared to physical circumferences and other measures. (18) Thus, this methodological approach may be a potential solution to collecting large scale anthropometric information and improve access to health information for those with monetary or geographic limitations. However, the consumer-level and accessible nature of this method fosters continual industry competition and thus, there are several mobile applications that claim to produce accurate automated anthropometric evaluations from smartphone-based imaging. As such, there are currently no studies, to our knowledge, that have fully assessed the equivalence of clinically significant measures of abdominal adiposity such as WC, hip circumference (HC), waist-to-hip ratio (WHR), and waist-to-height ratio (W:HT) produced by multiple mobile applications across smartphone types to a reference measurement. Therefore, the purpose of this study was to determine the agreement and precision of automated anthropometric assessments produced by three-dimensional mobile scanning applications compared to a reference tape measure.

METHODS

Participants

A total of 115 individuals (F: 69, M: 46) between ages 18 and 75 years were prospectively recruited for this cross-sectional study. Participants were excluded if they were younger than 18 or older than 75; were missing any limbs or part of a limb that influenced an accurate assessment of the primary anthropometric measures; were pregnant; trying to become pregnant; or breast feeding or lactating. The study took place from March 2022 through July 2022 and was conducted according to the guidelines laid down in the Declaration of Helsinki, with all procedures involving human participants approved by the university ethics committee (IRB#21-213). Written informed consent was obtained from all participants.

Procedures

Our procedures for visual body composition scanning have been previously reported (19) but are summarized below. Participants reported to the laboratory after abstention from food, beverages, supplements/medication, and exercise for ≥8 h. Upon arrival participants were asked to remove any external accessories (jewelry, shoes, etc.) and/or loose clothing and underwent measurements of height collected by a digital stadiometer (SECA, Hamburg, Germany), weight collected by a calibrated digital scale (SECA, Hamburg, Germany), and WC and HC collected using an aluminum tape measure. Following tape measurements, participants were lead to a specific area of the laboratory to complete the smartphone-based assessments. For scanning on each mobile application, participants were instructed to wear minimal form-fitting clothing. For example, female participants were instructed to wear a sports bra and tight-fitting shorts/leggings and male participants were instructed to wear compression shorts/tights only. Higher-waisted shorts that covered the participants bellybutton were altered to expose the participants entire abdominal region to the smartphone camera. Participants with long hair were instructed to tie their hair up so that no hair was present below the shoulder line.

Reference Tape Measurements

Traditional WC and HC were collected by an aluminum tape measure and these measurements were used as, or used to calculate, all reference variables for this study. Because there are no standardized measurement sites for WC and HC across mobile applications, estimates of WC and HC from these applications could be generated from different locations at or around the waist and hip regions. Therefore, the reference tape measurements were standardized to specific locations, where the reference WC was measured at the level of the iliac crest (20) and the reference HC was measured at the widest portion of the lateral hips (9) given the ease in the visual detection of pronounced or distinct body areas during the landmarking procedures of smartphone-based imaging. In addition, the American Heart Association waist circumference risk classifications are based off measurements collected at the level of the iliac crest. (21) All tape measurements were conducted by the same two investigators for all participants. WC and HC were used to produce a WHR by dividing WC by HC and measurements of W:HT by dividing WC by height collected from the digital stadiometer. For WHR measurements, the first WC was divided by the first HC and the second WC by the second

HC. Because height was measured a single timepoint, both the first and second WC were divided by the same height measurement. All measurements were conducted in duplicate and averaged to produce a final estimate.

Mobile Applications and Smartphone Types

Our procedures for each smartphone application and type have been described elsewhere (19). Two mobile applications were used for this study which included MeThreeSixty® (ME360; Size Stream LLC, Cary, NC, USA) and myBVI® (Select Research LTD, Malvern, England). Because mobile applications are frequently updated under real-world circumstances, and because these updates are often necessary to fix unavoidable issues with the application's performance, applications were updated daily prior to testing, when available. The study began with software version 3.3.0 (ME360$_{Apple}$), 3.2.2 (ME360$_{Samsung)}$, and 3.0.0 (myBVI®) and ended using versions 3.4.2 (both ME360) and 3.1.2 (myBVI®), respectively. Further, two different smartphones were used for this study to compare the precision and agreement of the applications across smartphone types. The smartphones used in this study included an iPhone® 12 Pro (Apple® Inc., Cupertino, CA, USA) and a Samsung Galaxy® S21+ (Samsung® Group, Suwon, South Korea). Assessments using the iPhone® were conducted using the same software version for the entirety of the study (iOS 15.0.1) but due to Samsung's® forced security updates, multiple software versions were employed for this smartphone specifically (One UI version 3.1, 4.0, and 4.1 and Android® version 11 and 12). All images for ME360 were collected using the front facing camera from both smartphones whereas images from myBVI® were collected using the front facing camera from the iPhone® only due to compatibility issues between myBVI® and Samsung Galaxy®.

Anthropometric Assessment Protocol for the Mobile Applications

Our procedures for collecting smartphone-based body composition estimates have been previously reported (19). To perform the assessments for each mobile application, participants were taken to a designated area of the laboratory that had did not have any objects or light behind the participants' back. All images were taken in front of a gray vinyl wall in this designated area and all external windows were covered so that no other background or external light source polluted the scanning region. The smartphone was positioned at a standardized distance from the

participant's mid-foot unless instructed otherwise by the application and a standardized height for all participants using a stationary tripod with adjustable angle settings. The smartphone order was assigned randomly for each participant. Each participant's personal information (age, sex, height, and weight) was uploaded into the application before testing commenced. The smartphone was locked into place at an angle determined appropriate by the mobile application. Once the smartphone and the participant were situated appropriately, participants were asked to stand in two positions, in accordance with the manufacturer guidelines, while images were collected. For the first image, participants faced the camera and stood with arms and feet positioned away from the torso. For the second image, participants were instructed to turn to their profile with either their left (ME360) or right (myBVI$^®$) shoulder facing the camera, face forward, extend the elbows completely, and place their hands against their lateral thigh while images were collected. All assessments were conducted in duplicate and subjectively inspected for quality to ensure there were no errors were apparent during landmarking procedures. WC and HC from ME360 were provided directly by the mobile application. WHR$_{ME360}$ was calculated as WC$_{ME360}$ divided by HC$_{ME360}$. Because myBVI$^®$ provides WHR and W:HT, but not WC or HC, WC$_{myBVI}$ was calculated as height from the stadiometer multiplied by W:HT$_{myBVI}$. The newly produced WC$_{myBVI}$ was divided by WHR$_{myBVI}$ to produce HC$_{myBVI}$. Because myBVI$^®$ requires the user to round their height, and because ME360$_{W:HT}$ used height measured by stadiometer, we calculated W:HT$_{myBVI}$ by dividing WC$_{myBVI}$ from the stadiometer height. For WHR measurements from each application, the first automated WC was divided by the first automated HC and the same was done for the second scans. Similar to the reference method, both the first and second automated WC were divided by the same height measurement. All measurements were conducted in duplicate and averaged to produce a final estimate.

Statistical Analysis

We conducted a non-directional power analysis to determine the sample size necessary to detect significant differences using a paired-samples t-test (the primary statistical analysis for determining group mean differences). Prior to analysis we determined ±4.0 cm to be a meaningful mean difference (MD) and thus, using a MD of 4.0 ± 6.0 and an $\alpha = 0.05$ it was determined that 20 participants were necessary to observe at least 80% power. All outcome

variables were normally distributed as assessed by Shapiro-Wilk and visual inspection of Q-Q plots. Means and 95%CI for each device were calculated for WC, HC, WHR, and W:HT and the MD and 95%CI for each variable were calculated as the mobile application in question minus the reference. Test-retest precision of each anthropometric assessment were assessed using intraclass correlation coefficients (ICC) with two-way, random effects, and absolute agreement. Precision was also measured using precision error (PE) and root mean square coefficient of variation (RMS-%CV). For ME360, precision metrics were used to determine precision between smartphone types using the first scan from each smartphone. Because WHR and W:HT are unitless, PE was not calculated for these variables. A device error resulted in one participant (n=1) missing a single scan from one smartphone (Samsung®) for ME360. Therefore, 114 participants were used to determine precision and 115 participants were used to assess agreement. An average of the two scans for each mobile application were used to determine agreement and only one scan was used for the participant with a missing scan. Equivalence testing was used to determine equivalence with the reference tape measurements for WC, HC, WHR, and W:HT using 5% equivalence regions. Additionally, and because WC is often used to assess abdominal obesity for the evaluation of cardiometabolic health risk, the percentage of correct abdominal obesity classifications according to the guidelines put forth by the American Heart Association ($\geq$ 88 cm for females; $\geq$ 102 cm for males) are presented for each mobile assessment. (21) Agreement with the reference method was also assessed by separate paired samples t test, Pearson correlation coefficients, root mean square error (RMSE=$\sqrt{\sum}$(predicted-actual)^2/n) and standard error of the estimate (SEE). Individual accuracy was assessed using the methods of Bland and Altman (22) to determine the 95% limits of agreement (LOA) and regression techniques were used to determine proportional biases. Subgroup analyses were conducted for sex and racial differences. For race, analyses were conducted for non-Hispanic white and non-Hispanic Black/African-American (B/AA) individuals (**Table 1**). Data from other racial and ethnic groups were included in the complete sample analyses only. Anthropometric differences between groups were assessed by independent samples t-test. Statistical significance was accepted at $p < 0.05$. Given the large sample size and observational nature of the study, a power analysis was not conducted. Data was analyzed using the TOSTER package (23) in R version 4.1.2, IBM SPSS version 27, and Microsoft Excel version 16.

RESULTS

Participant characteristics are reported as mean and the 95% confidence interval (95%CI) or as a percentage of the total for each column (**Table 1**).

Precision Analysis

Results of the precision analysis are presented in **Table 2**. For precision, all ICCs, including assessments between smartphone types, ranged from 0.933 to 0.998 (all $p < 0.001$). For WC and HC, $ME360_{Apple}$ produced the lowest PE and RMS-%CV followed by tape measurements and $ME360_{Samsung}$. myBVI$^®$ had the largest PE and RMS-%CV for both WC and HC across applications. Overall, precision was lower between smartphone types for WC and HC produced by $ME360_{Apple \ x \ Samsung}$. RMS-%CV for WHR was lowest for $ME360_{Apple}$ (0.70%) which was slightly lower than $ME360_{Samsung}$ (0.80%), but both $ME360_{WHR}$ estimates had lower RMS-%CV than those conducted by tape measurement (1.05%). myBVI$^®$ had the highest RMS-%CV for WHR (2.43%) which was also higher than the RMS-%CV between smartphone types for $ME360_{Apple \ x \ Samsung}$ (1.98%).

Agreement Analysis

Results of the agreement analysis for the total sample are presented in **Table 3.** Overall, the number of correct abdominal obesity classifications for each mobile application were 103 (89.6%) for $ME360_{Apple}$, 104 (90.4%) for $ME360_{Samsung}$, and 106 (92.2%) for myBVI$^®$. Of the incorrect classifications, both $ME360_{Apple}$ (12 incorrect) and myBVI$^®$ (9 incorrect) incorrectly classified participants as having abdominal obesity on three occasions with the remaining incorrectly classifying participants as not having abdominal obesity. $ME360_{Samsung}$ incorrectly classified participants (11 incorrect) as having abdominal obesity on five occasions with the remaining incorrectly classifying participants as having abdominal obesity.

Paired t-tests revealed that all variables produced by the mobile applications differed significantly from the reference tape measurement for the total sample. WC was slightly underestimated by each mobile application (all MD: $\leq$ -2.5 cm, $p < 0.050$) relative to the reference method for each application. Conversely, HC was slightly overestimated across each mobile application relative to the reference method with myBVI$^®$ demonstrating the lowest MD (1.9 cm) followed by ME360 with negligible differences between smartphone types (all $p <$

0.050). MD for WHR and W:HT were similar across devices. Despite significantly different MD, equivalence testing revealed that all devices demonstrated equivalence to the reference method using a 5% equivalence region for WC, HC, and W:HT (**Figure 1, Table 3**). The slight underestimation of WC and the slight overestimation of HC resulted in a non-significant equivalence test for WHR. ME360 demonstrated the highest RMSE for WC (±6.5 cm for each) but the lowest RMSE for HC (±4.7 to ±5.1 cm). RMSE for WHR and W:HT were similar across applications.

Results for the Bland-Altman analysis for the total sample are displayed in **Table 3** and illustrated in **Figure 2**. LOA ranged from ±9.0 to ±12.4 cm for WC and from ±6.7 to ±10.7 cm for HC. For WHR and W:HT, LOA were similar across applications. Significant proportional bias was observed for WC estimates produced by ME360 but not myBVI®. Conversely, significant proportional bias was observed for HC estimates produced by all methods.

Sex Differences in Agreement

Agreement analysis by sex groups are presented in **Table 4**. When stratified by sex, there were no significant differences between the WC estimates produced by each device and the reference method for males (all $p > 0.050$); however, WC was significantly underestimated for females across all devices (all $p < 0.001$). Significant overestimations of HC were observed for both males and females using ME360 ($p < 0.001$) but only for males using myBVI® ($p < 0.001$). WHR was significantly underestimated for both males and females across all applications (all $p < 0.05$). For WC, RMSE were lower for males across all devices. RMSE for HC estimates were negligible between males and females using ME360 but were substantially higher for males using myBVI®. RMSE for WHR were similar between males and females other than WHR estimates produced by ME360$_{Apple}$ which, for females, which was more than double that of males.

Results for the Bland-Altman analysis by sex group are displayed in **Table 4**. LOA ranged from ±6.6 to ±12.9 cm for WC and were smaller in males across all devices. LOA ranged from ±6.5 to ±10.1 cm for HC and were smaller in females across all devices, albeit similar for ME360. For WHR and W:HT, LOA were similar between males and females although all LOA were slightly lower for males. Significant proportional biases were observed for WC estimates produced by ME360 (all $p < 0.001$), but not myBVI® (both $p > 0.050$) and were similar between

males and females. Significant proportional biases were observed for HC estimates produced by myBVI® and were similar between males and females (both p < 0.050). For HC produced by ME360, no proportional bias was observed for males, but was observed for females using ME360$_{Apple}$ (p < 0.010) but not ME360$_{Samsung}$ (p = 0.058). Significant proportional biases were observed for WHR produced by ME360 in both males and females (all p < 0.050) but not for WHR produced by myBVI® (p < 0.050).

Racial Differences in Agreement

Agreement analysis by the race groups evaluated in this study are presented in **Table 5**. WC estimates from ME360$_{Apple}$, but not ME360$_{Samsung}$, were significantly underestimated for both White and B/AA participants (p < 0.010). Although non-significant, the MD for WC produced by ME360$_{Samsung}$ for B/AA participants was more than double the MD for White participants. myBVI® significantly underestimated WC for White participants (p < 0.001) but not B/AA participants (p = 0.08). HC estimates from each application were all significantly overestimated (all p < 0.050); however, MD for HC estimates were markedly higher for B/AA participants compared to White participants. WHR estimates were significantly underestimated for all applications (all p < 0.001) and were similar between White and B/AA participants. For WC, RMSE were similar between White and B/AA participants across devices; however, RMSE for HC were substantially higher for B/AA participants across applications. RMSE for WHR and W:HT were similar between White and B/AA participants across applications.

Results for the Bland-Altman analysis by race are displayed in **Table 5**. LOA ranged from ±9.0 to ±13.4 cm for WC and were similar for White and B/AA participants. For HC, LOA ranged from ±5.5 to ±11.9 and were noticeably higher for B/AA participants for each device. For WHR and W:HT, LOA were similar for White and B/AA participants. Significant proportional bias was observed for WC and W:HT estimates produced by ME360 (all p < 0.050) but not myBVI® which were similar for White and B/AA participants. For HC, significant proportional bias was observed across all devices in White participants (p < 0.010) but only for myBVI® in B/AA participants (p = 0.024); however, coefficients for each device were similar between groups. There was no significant proportional bias observed for WHR estimates across devices,

although the coefficients for WHR produced by ME360 were non-existent for White participants and high for B/AA participants.

DISCUSSION

While traditional tape measurements are considered to be cost effective and accessible, there are questions regarding their social acceptance and reliability, (24) particularly for those with overweight or obesity. (25) As such, this study sought comprehensively evaluate the precision and agreement of automated and clinically significant anthropometric variables across multiple mobile applications, and smartphones, against a reference tape measurement. The principle findings were 1) all variables produced by each mobile application and between smartphones exhibited acceptable precision which were comparable with tape measurements; 2) WC, HC, and W:HT from all mobile applications demonstrated equivalence with tape measurements; however, WC was slightly underestimated whereas HC was slightly overestimated; 3) the slight under- and overestimations for WC and HC were small enough to demonstrate equivalence, but resulted in non-equivalence for WHR across all automated methods; 4) there were no differences between WC produced by the automated methods and the reference in males, but WC was significantly underestimated across all applications in females; and 5) some variation existed across applications, but all variables demonstrated slightly lower agreement for B/AA participants compared to White participants which may be a product of the weight status differences between groups or demographics of the populations used for method development. Overall, the results of our study support the use of WC and HC estimates produced from automated mobile applications, but demonstrates the importance of accurate automation for WC and HC estimates given their influence on other anthropometric assessments and clinical health markers.

First, each mobile application used in our study demonstrated acceptable precision for all automated assessments. There was a slight drop-off in precision when estimates were compared between smartphone types; however, precision remained within an acceptable range. While there are only a few studies have evaluated the precision of automated anthropometrics using mobile applications, (16,26) there is contention regarding the precision of traditional tape measurements citing inaccuracies between self-measured and professionally measured assessments (27) and measurements for those of higher weight status. (27,28) Another method shown to produce

reliable digital anthropometrics is three-dimensional (3D) optical scanning. (29) However, three-dimensional scanners are generally unavailable for simple but important circumference measures in clinical practice, especially for those in lower SES or rural areas. Thus, the limitations of other methods highlight the need for precise estimates that are both accessible and cost-effective. In our study, all automated anthropometrics produced precision estimates that were similar to the tape measurements performed in our study (which were conducted professionally) and those produced by three-dimensional scanning in others. (9,29) Interestingly, the ME360$_{Apple}$ application produced better precision estimates than those collected via tape measure with ME360$_{Samsung}$ only marginally lower. It is possible that these small differences are due to differences in the developmental software (i.e., iOS or Android®), where initial mobile applications are built for a specific device (i.e., iPhone® or Samsung®) with operating systems that require a particular coding language. (30) Despite this, all ICCs were > 0.930 and PE were all ≤ 2.0 cm. Considering that this degree of precision was produced simply from two two-dimensional pictures on highly accessible mobile applications, it is plausible that these applications be considered reliable and comparable to traditional methods.

Several investigations have evaluated the agreement between automated anthropometrics from a mobile application and traditional tape measurements and have demonstrated considerable variation, (16,26,31,32) likely due to the different mobile applications employed in each study. Overall, WC, HC, and W:HT across all mobile applications in our study demonstrated equivalence to our reference method; however, there were slightly significant under- and overestimations for WC and HC, respectively, for each application. These bi-directional biases, where an underestimated WC was divided by a larger overestimated HC, resulted in significant underestimation of WHR for each application that did not demonstrate equivalence. So, while biases for WC and HC were relatively small, these small differences manifested in discrepancies across other variables. This relationship is also reflected in our results for W:HT, where all mobile applications underestimated W:HT as an extension of an underestimated WC. These small inconsistencies for WC and HC are problematic considering that WC and HC estimates are commonly used in other anthropometric screening tools to predict several health risks. (10,33) This is especially concerning considering the proportional biases and large LOA observed in our study. Specifically, we found that both ME360 applications demonstrated significant proportional bias for WC, where WC was underestimated to a greater

degree in those with larger average WC. Significant proportional bias was observed for HC produced by both ME360 applications, but these biases were relatively small and much smaller than the proportional biases produced for WC using the same application. Because participants with a higher average WC had greater underestimations of WC without similar underestimations of HC, WHR was underestimated using this application. Interestingly, there was no proportional bias for WC using myBVI®; however, HC produced by myBVI® demonstrated significant proportional bias, where HC was overestimated for those at larger average HC. Similar to ME360, albeit in differing directions, the overestimation of HC without simultaneous overestimations in WC led to underestimations of WHR. Therefore, while automated WC and HC estimates from a mobile application may demonstrate equivalence, those planning to employ these estimates as a part of a larger screening battery should do so with caution; although WC estimates produced by each mobile application did show utility in correctly determining abdominal obesity classification; a common cardiometabolic health risk assessment.

There are several issues that may explain our aforementioned results. First, the artificial intelligence used to develop each application is dependent upon the method used to train the application. For instance, if the mobile application was trained by a 3D scanner it is possible that comparisons to a tape measurement would result in lower levels of agreement. However, many mobile applications are trained by both 3D scanning and traditional tape measures, and recent investigations show agreement in body circumferences assessed by tape measurements, 3D scanners, and mobile applications. (16) Typically, these studies determine agreement by comparing automated measures to tape measurements taken at sites specifically defined by the mobile application. (16,31) While this methodological approach may result in better agreement, it may also limit real-world application given that self- or professionally-measured circumferences may be taken from considerably different locations than from those suggested by the mobile application. Moreover, measurement sites may be markedly different between mobile applications making them difficult to compare. Therefore, to determine their "real-world" performance we standardized the location of each tape measurement. It should be noted, however, that while the automated WC and HC were equivalent to tape measurements, this is specific to the performance of our investigators and the location in which our tape measurements were taken and thus, estimates taken by individual users, by other professionals, or at different locations may lead to differing results.

Accepted manuscript

In addition to what has been previously suggested, our results may also be explained by our sex and race comparisons. For males, there were no significant differences between the WC measures and the reference for any mobile application. In fact, MD in WC for males were all ≤ -0.6 cm which is comparable to the results for both males and females produced by Nana et al. (31) using a single mobile application. Conversely, WC was significantly underestimated by ≥ -3.0 cm across applications for females. Size differences between our female participants and those in the study by Nana et al. (31) may explain the differences between studies, where our female participants had substantially higher weight, WC, and HC. However, male participants in our study were also much larger. As previously suggested by Neufeld et al. (32), it is possible that differences in clothing between male and female participants, combined with the larger size of our participants, may explain our findings. Specifically, male participants wore only compression shorts/tights whereas female participants wore tights and a sports bra. It is common for females to wear high-waisted tights that cover a significant area (and often the majority) of the abdomen where the automated WC would be collected. Although steps were taken to alleviate this in our study, the automated landmarking procedures used by each mobile application may be altered when a considerable amount of abdominal mass is covered. Additionally, we observed that clothing worn by females was often tighter fitting than clothing worn by males. The tightness of the clothing, especially in female participants of larger body size, may alter the natural body shape in areas around the abdomen. It is also possible that the discrepancies in WC between the automated assessments and the reference in females was due to the inherent difficulty in measuring the female waist. As you descend the body's vertical axis, the variation in WC along that axis is more exaggerated in females compared to males as a result of normal adult female body shape. (34,35) Thus, it is possible that these issues can introduce error into the automated measurement process as each application attempts to detect specific landmarks for its WC estimate.

The sex differences in HC for myBVI® may also explain a number of our other findings. For males, myBVI® significantly overestimated HC which was not observed for females. Interestingly, the majority of our male participants had either overweight or obesity, in addition to larger WC and HC, compared to our female participants. Considering that the proportional biases for HC from myBVI® were nearly identical between sexes, it is more likely that the larger WC and weight status contributed to these differences. It is well known that males tend to

deposit more fat in the anterior abdominal area whereas females tend to deposit more fat in the gluteal-femoral region (36) leading to a larger WC for males. At higher degrees of overweight and obesity, the excessive accumulation of fat in the abdominal area is exposed to gravity and may extend well into the pubic region. Given that the pubic area is within the normal assessment region for HC, it is possible that the larger abdominal fat mass in male participants extended into the HC area for the automated assessments whereas tape measurements could avoid this interference. This potential discrepancy can also be observed in our comparisons by race. B/AA participants had significantly higher weight, WC, and HC compared to their White counterparts and 50% of our B/AA participants had obesity, with all but one of our participants with severe obesity being B/AA males (the one participant was a B/AA female). Further, HC was overestimated to a greater degree for B/AA participants but without differences in proportional biases. Given the large differences in weight and WC and the lack of proportional bias in HC in this group, it is possible that at the extremes of abdominal adiposity, abdominal fat mass may impede the anterior area of the HC measurement resulting in an automated HC measurement that accounts for a portion of the waist, leading to HC overestimation. The potential impedance of the abdomen in larger individuals, coupled with the intrinsically larger distributions of fat in the gynoid region for B/AA individuals, (37) could result in larger measurement errors that may be exaggerated for B/AA males.

In conclusion, automated anthropometrics produced by two dimensional pictures from mobile applications are cost-effective and accessible tools that can be used to collect clinically significant anthropometric information. Automated anthropometrics from mobile applications demonstrate high levels of precision regardless of the smartphone used. Mobile anthropometrics also demonstrate agreement with traditional tape measurements (via equivalence testing) for WC and HC estimates. However, slight deviations in WC and HC, which may be due to technical issues that are exaggerated in estimates for female and B/AA participants, may lead to inaccuracies for other measurements such as WHR. As such, these data support the overall use of this technique for estimates of WC and HC, but individuals should include these data in comprehensive health assessments with caution given the range of individual error and over-and-underestimations. Despite this, WC estimates produced by each mobile application demonstrate utility in assessing abdominal obesity. These findings also support the potential use of this assessment method in prospective studies, where participants could self-assess across an

intervention without the need for additional laboratory visits. However, future research examining the accuracy of self- and at-home assessments are necessary in addition to studies examining accurate assessment over time.

FUNDING

CONFLICTS OF INTEREST

The authors declare no conflicts of interest. GMT has received in-kind donations of 3-dimensional optical imaging equipment through research contracts with SizeStream, LLC; Naked Labs, Inc.; and Prism Labs, Inc. These entities played no role in the present study.

AUTHOR CONTRIBUTIONS

AJG: Conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, roles/writing – original draft.

CFB: Data curation, investigation, methodology, validation, writing – review & editing

GMT: Conceptualization, formal analysis, investigation, methodology, validation, visualization, writing – review & editing.

Accepted manuscript

REFERENCES

1. Anekwe CV, Jarrell AR, Townsend MJ, Gaudier GI, Hiserodt JM, Cody Stanford F. Socioeconomics of Obesity. Curr Obes Rep. 2020 Sep;9(3):272–9.

2. Okobi OE, Ajayi OO, Okobi TJ, Anaya IC, Fasehun OO, Diala CS, et al. The Burden of Obesity in the Rural Adult Population of America. Cureus. 2021;13(6):e15770.

3. Bixby H, Bentham J, Zhou B, Di Cesare M, Paciorek CJ, Bennett JE, et al. Rising rural body-mass index is the main driver of the global obesity epidemic in adults. Nature. 2019 May;569(7755):260–4.

4. Volaco A, Cavalcanti AM, Filho RP, Précoma DB. Socioeconomic Status: The Missing Link Between Obesity and Diabetes Mellitus? Curr Diabetes Rev. 2018;14(4):321–6.

5. Setiono FJ, Guerra LA, Leung C, Leak TM. Sociodemographic characteristics are associated with prevalence of high-risk waist circumference and high-risk waist-to-height ratio in U.S. adolescents. BMC Pediatr. 2021 May 3;21(1):215.

6. Pujilestari CU, Nyström L, Norberg M, Weinehall L, Hakimi M, Ng N. Socioeconomic inequality in abdominal obesity among older people in Purworejo District, Central Java, Indonesia – a decomposition analysis approach. Int J Equity Health. 2017 Dec 12;16:214.

7. Janssen I, Katzmarzyk PT, Ross R. Waist circumference and not body mass index explains obesity-related health risk. Am J Clin Nutr. 2004 Mar;79(3):379–84.

8. Abe M, Fujii H, Funakoshi S, Satoh A, Kawazoe M, Maeda T, et al. Comparison of Body Mass Index and Waist Circumference in the Prediction of Diabetes: A Retrospective Longitudinal Study. Diabetes Ther. 2021 Oct 1;12(10):2663–76.

9. Jaeschke L, Steinbrecher A, Pischon T. Measurement of Waist and Hip Circumference with a Body Surface Scanner: Feasibility, Validity, Reliability, and Correlations with Markers of the Metabolic Syndrome. PLoS ONE. 2015 Mar 6;10(3):e0119430.

Accepted manuscript

10. Price GM, Uauy R, Breeze E, Bulpitt CJ, Fletcher AE. Weight, shape, and mortality risk in older persons: elevated waist-hip ratio, not high body mass index, is associated with a greater risk of death. Am J Clin Nutr. 2006 Aug;84(2):449–60.

11. Record shares of Americans have smartphones, home broadband [Internet]. Pew Research Center. [cited 2021 Sep 8]. Available from: https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/

12. Vangeepuram N, Mayer V, Fei K, Hanlen-Rosado E, Andrade C, Wright S, et al. Smartphone ownership and perspectives on health apps among a vulnerable population in East Harlem, New York. mHealth. 2018 Aug 8;4:31.

13. McCloskey ML, Thompson DA, Chamberlin B, Clark L, Johnson SL, Bellows LL. Mobile Device Use Among Rural, Low-Income Families and the Feasibility of an App to Encourage Preschoolers' Physical Activity: Qualitative Study. JMIR Pediatr Parent. 2018 Dec 6;1(2):e10858.

14. Batsis JA, McClure AC, Weintraub AB, Kotz DF, Rotenberg S, Cook SB, et al. Feasibility and acceptability of a rural, pragmatic, telemedicine-delivered healthy lifestyle programme. Obes Sci Pract. 2019;5(6):521–30.

15. Xiao H, Barber JP. The Effect of Perceived Health Status on Patient Satisfaction. Value Health. 2008 Jul 1;11(4):719–25.

16. Smith B, McCarthy C, Dechenaud ME, Wong MC, Shepherd J, Heymsfield SB. Anthropometric evaluation of a 3D scanning mobile application. Obesity. 2022;30(6):1181–8.

17. Graybeal AJ, Brandner CF, Tinsley GM. Validity and reliability of a mobile digital imaging analysis trained by a four-compartment model. J Hum Nutr Diet. 2022; 1–7.

18. Graybeal AJ, Brandner CF, Tinsley GM, Haynes H, Stavres J. Associations between visceral adipose tissue estimates produced by near-infrared spectroscopy, mobile anthropometrics, and traditional body composition assessments and estimates derived from dual-energy X-ray absorptiometry. Brit J Nutr. 2022; 1-11.

19. Graybeal AJ, Brandner CF, Tinsley GM. Visual body composition assessment methods: a 4-compartment model comparison of smartphone-based artificial intelligence for body composition estimation in healthy adults. Clin Nutr. 2022;41(11):2464-72.

20. Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. Nat Rev Endocrinol. 2020;16(3):177–89.

21. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al; American Heart Association; National Heart, Lung, and Blood Institute. Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation. 2005;112(17):2735-52.

22. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. The Lancet. 1986 Feb 8;327(8476):307–10.

23. Lakens D. Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. Soc Psychol Personal Sci. 2017 May;8(4):355–62.

24. Verweij LM, Terwee CB, Proper KI, Hulshof CT, Mechelen W van. Measurement error of waist circumference: gaps in knowledge. Public Health Nutr. 2013 Feb;16(2):281–8.

25. Minetto MA, Pietrobelli A, Busso C, Bennett JP, Ferraris A, Shepherd JA, et al. Digital Anthropometry for Body Circumference Measurements: European Phenotypic Variations throughout the Decades. J Pers Med. 2022 Jun;12(6):906.

26. Macdonald EZ, Vehrs PR, Fellingham GW, Eggett D, George JD, Hager R. Validity and Reliability of Assessing Body Composition Using a Mobile Application. Med Sci Sports Exerc. 2017 Dec;49(12):2593–9.

27. Carranza Leon BG, Jensen MD, Hartman JJ, Jensen TB. Self-Measured vs Professionally Measured Waist Circumference. Ann Fam Med. 2016 May;14(3):262–6.

28. Contardo Ayala AM, Nijpels G, Lakerveld J. Validity of self-measured waist circumference in adults at risk of type 2 diabetes and cardiovascular disease. BMC Med. 2014 Oct 2;12:170.

29. Rumbo-Rodríguez L, Sánchez-SanSegundo M, Ferrer-Cascales R, García-D'Urso N, Hurtado-Sánchez JA, Zaragoza-Martí A. Comparison of Body Scanner and Manual Anthropometric Measurements of Body Shape: A Systematic Review. Int J Environ Res Public Health. 2021 Jun 8;18(12):6213.

30. White J. Going native (or not): Five questions to ask mobile application developers. Australas Med J. 2013 Jan 31;6(1):7–14.

31. Nana A, Staynor JMD, Arlai S, El-Sallam A, Dhungel N, Smith MK. Agreement of anthropometric and body composition measures predicted from 2D smartphone images and body impedance scales with criterion methods. Obes Res Clin Pract. 2022 Jan 1;16(1):37–43.

32. Neufeld EV, Seltzer RA, Sazzad T, Dolezal BA. A Multidomain Approach to Assessing the Convergent and Concurrent Validity of a Mobile Application When Compared to Conventional Methods of Determining Body Composition. Sensors. 2020 Oct 29;20(21):6165.

33. Yoo EG. Waist-to-height ratio as a screening tool for obesity and cardiometabolic risk. Korean J Pediatr. 2016 Nov;59(11):425–31.

34. Wang J, Thornton JC, Bari S, Williamson B, Gallagher D, Heymsfield SB, et al. Comparisons of waist circumferences measured at 4 sites. Am J Clin Nutr. 2003 Feb 1;77(2):379–84.

35. Bosy-Westphal A, Booke CA, Blöcker T, Kossel E, Goele K, Later W, et al. Measurement Site for Waist Circumference Affects Its Accuracy As an Index of Visceral and Abdominal Subcutaneous Fat in a Caucasian Population. J Nutr. 2010 May 1;140(5):954–61.

36. Nauli AM, Matin S. Why Do Men Accumulate Abdominal Visceral Fat? Front Physiol [Internet]. 2019 [cited 2022 Jul 1];10. Available from: https://www.frontiersin.org/article/10.3389/fphys.2019.01486

37. Stults-Kolehmainen MA, Stanforth PR, Bartholomew JB. Fat in Android, Trunk, and Peripheral Regions Varies by Ethnicity and Race in College Aged Women. Obesity. 2012;20(3):660–5.

Accepted manuscript

Table 1. Participant Characteristics

N=115		Total	Male	Female
Sex			46 (40.0%)	69 (60.0%)
Race	White	83 (72.2%)	31 (67.4%)	52 (75.4%)
	Black/AA	28 (24.3%)	14 (30.4%)	14 (20.3%)
	Asian	4 (3.5%)	1 (2.2%)	3 (4.3%)
Ethnicity	Hispanic	7 (6.1%)	2 (4.3%)	5 (7.2%)
Anthropometry	Age (y)	29.4 (27.1, 31.7)	26.5 (23.8, 29.1)	31.3 (28.0, 34.6)[c]
	Height (cm)	170.2 (168.4, 172.0)	178.0 (175.5, 180.4)	165.0 (163.5, 166.5)[a]
	Weight (kg)	80.4 (76.3, 84.4)	94.3 (87.4, 101.1)	71.07 (67.5, 74.7)[a]
	BMI (kg/m^2)	27.5 (26.2, 28.7)	30.2 (27.9, 32.5)	25.6 (24.4, 26.9)[b]
	Waist (cm)	91.4 (88.5, 94.3)	96.2 (91.2, 101.3)	88.2 (84.9, 91.6)[b]
	Hip (cm)	103.0 (101.7, 106.1)	106.2 (102.3, 110.5)	102.4 (99.8, 105.0)
	Waist:Hip	0.88 (0.86, 0.89)	0.90 (0.89, 0.92)	0.86 (0.84, 0.88)[b]
	Waist:Height	0.54 (0.52, 0.56)	0.54 (0.51, 0.57)	0.54 (0.51, 0.57)
BMI Classification	≥ 40 kg/m^2	5 (4.3%)	4 (8.7%)	1 (1.4%)
	30 – 39.9 kg/m^2	23 (20.0%)	11 (23.9%)	12 (17.4%)
	25 – 29.9 kg/m^2	39 (33.9%)	20 (43.5%)	19 (27.5%)
	< 25 kg/m^2	48 (41.7%)	11 (23.9%)	37 (53.6%)
N=104		**White (n=76)**	**Black (n=28)**	
Anthropometrics	Height (cm)	170.5 (168.4, 172.6)	171.2 (167.0, 175.5)	
	Weight (kg)	77.1 (73.4, 80.8)	94.1 (82.0, 106.1)[c]	
	BMI (kg/m^2)	26.3 (25.2, 27.4)	31.8 (28.0, 35.7)[b]	
	Waist (cm)	89.4 (86.6, 92.3)	99.9 (91.2, 108.6)[c]	
	Hip (cm)	102.5 (100.3, 104.7)	109.5 (103.1, 116.1)[c]	
	Waist:Hip	0.87 (0.86, 0.89)	0.91 (0.87, 0.94)[c]	
	Waist:Height	0.53 (0.51, 0.54)	0.59 (0.53, 0.64)[c]	
BMI Classification	≥ 40 kg/m^2	0 (0.0%)	5 (17.9%)	
	30 – 39.9 kg/m^2	14 (18.4%)	9 (32.1%)	
	25 – 29.9 kg/m^2	31 (40.8%)	4 (14.3%)	
	< 25 kg/m^2	31 (40.8%)	10 (35.7%)	

AA, African-American; BMI, body mass index. [a] statistically significant at p <0.001; [b] statistically significant at p <0.010; [c] statistically significant at p <0.050

Accepted manuscript

Table 2. Precision analysis of smartphone-based automated anthropometrics

N=114		ICC[a]	RMS-%CV	PE[b]
WC	Tape Measure	0.998	0.78	0.63
	ME360$_{Apple}$	0.998	0.69	0.54
	ME360$_{Samsung}$	0.996	1.11	0.87
	myBVI	0.995	1.43	1.12
	ME360$_{Apple \, x \, Samsung}$	0.979	2.54	1.91
HC	Tape Measure	0.997	0.70	0.65
	ME360$_{Apple}$	0.998	0.54	0.51
	ME360$_{Samsung}$	0.994	0.95	0.90
	myBVI	0.982	2.07	1.91
	ME360$_{Apple \, x \, Samsung}$	0.979	1.72	1.59
WHR	Tape Measure	0.984	1.05	
	ME360$_{Apple}$	0.992	0.70	
	ME360$_{Samsung}$	0.991	0.80	
	myBVI	0.933	2.43	
	ME360$_{Apple \, x \, Samsung}$	0.952	1.98	
W:HT	Tape Measure/Stadiometer	0.997	0.79	
	ME360$_{Apple}$	0.997	0.71	
	ME360$_{Samsung}$	0.994	1.12	
	myBVI	0.995	1.42	
	ME360$_{Apple \, x \, Samsung}$	0.978	2.51	

HC: hip circumference; ICC: intraclass correlation coefficient (two-way random effects, absolute agreement, single measurement); PE: precision error; RMS-%CV: root mean square coefficient of variation (%); WC: waist circumference; WHR: waist-to-hip ratio; W:HT: waist-to-height ratio. [a] p < 0.001 for all measurements; [b] measurements are in cm

Table 3. Agreement between smartphone-based automated anthropometrics and reference tape measurements

N=115		Mean (95%CI)[a]	MD (95%CI)[a]	EQ	RMSE	95% LOA	SEE	β	R^2
WC	Tape Measure	91.4 (88.5, 94.4)							
	ME360_Apple	89.1 (86.6, 91.7)	-2.3 (-3.4, -1.2)[b]	Y	6.5	11.9	2.4	-0.15[b]	0.86
	ME360_Samsung	89.9 (87.4, 92.4)	-1.6 (-2.7, -0.4)[d]	Y	6.5	12.4	2.6	-0.15[b]	0.84
	myBVI	88.9 (85.9, 91.9)	-2.5 (-3.4, -1.7)[b]	Y	5.2	9.0	1.5	0.03	0.92
HC	Tape Measure	103.9 (101.7, 106.1)							
	ME360_Apple	107.0 (105.0, 109.1)	3.1 (2.5, 3.8)[b]	Y	4.7	6.8	1.4	-0.07[d]	0.91
	ME360_Samsung	107.6 (105.5, 109.7)	3.7 (3.1, 4.4)[b]	Y	5.1	6.7	1.5	-0.06[d]	0.92
	myBVI	105.8 (103.1, 108.4)	1.9 (0.87, 2.9)[b]	Y	5.8	10.7	2.1	0.20[b]	0.87
WHR	Tape Measure	0.88 (0.86, 0.89)							
	ME360_Apple	0.83 (0.82, 0.84)	-0.05 (-0.06, -0.04)[b]	N	0.07	0.11	0.05	-0.06	0.48
	ME360_Samsung	0.83 (0.82, 0.85)	-0.04 (-0.05, -0.03)[b]	N	0.07	0.11	0.05	-0.08	0.47
	myBVI	0.84 (0.83, 0.85)	-0.04 (-0.05, -0.03)[b]	N	0.06	0.09	0.03	-0.07	0.66
W:HT	Tape Measure/Stadiometer	0.54 (0.52, 0.56)							
	ME360_Apple	0.52 (0.51, 0.54)	-0.014 (-0.021, -0.008)[b]	Y	0.04	0.07	0.01	-0.17[d]	0.86
	ME360_Samsung	0.53 (0.51, 0.54)	-0.009 (-0.016, -0.002)[c]	Y	0.04	0.07	0.02	-0.18[d]	0.85
	myBVI	0.52 (0.51, 0.54)	-0.015 (-0.020, -0.010)[b]	Y	0.03	0.05	0.01	0.01	0.92

95%CI: 95% confidence interval; β: regression coefficient produced from linear regression as used to assess proportional bias; EQ: equivalence; HC: hip circumference; LOA: limits of agreement; MD: mean difference; RMSE: root mean square error; SEE: standard error of the estimate; WC: waist circumference; WHR: waist-to-hip ratio; W:HT: waist-to-height ratio. [a] Expressed as mean (95% confidence interval); [b] Indicates statistical significance at $p < 0.001$; [c] indicates statistical significance at $p < 0.010$; [d] indicates statistical significance at $p < 0.05$

Table 4. Agreement between smartphone-based automated anthropometrics and reference tape measurements by sex groups

N=115		Males (n=46)					Females (n=69)				
		MD (95%CI)[a]	RMSE	LOA	β	R^2	MD (95%CI)	RMSE	LOA	β	R^2
WC	ME360$_{Apple}$	-0.3 (-1.7, 1.2)	4.8	9.6	-0.20[b]	0.94	-3.7 (-5.2, -2.1)[b]	7.4	12.6	-0.22[b]	0.80
	ME360$_{Samsung}$	0.6 (-1.0, 2.2)	5.3	10.4	-0.22[b]	0.93	-3.0 (-4.6, -1.4)[b]	7.2	12.9	-0.21[b]	0.78
	myBVI	-0.4 (-1.3, 0.6)	3.3	6.6	-0.01	0.93	-4.0 (-5.1, -2.8)[b]	6.2	9.3	0.01	0.89
HC	ME360$_{Apple}$	2.6 (1.5, 3.7)[b]	4.6	7.2	-0.02	0.92	3.5 (2.7, 4.3)[b]	4.8	6.5	-0.11[c]	0.91
	ME360$_{Samsung}$	3.5 (2.4, 4.5)[b]	5.1	6.8	-0.03	0.93	3.9 (3.1, 4.7)[b]	5.2	6.7	-0.08	0.90
	myBVI	5.9 (4.4, 7.4)[b]	6.0	10.1	0.14[d]	0.89	-0.8 (-1.7, 0.1)	3.8	7.4	0.13[c]	0.91
W:Hip	ME360$_{Apple}$	-0.02 (-0.03, -0.01)[c]	0.03	0.08	-0.30[d]	0.51	-0.06 (-0.08, -0.05)[b]	0.07	0.11	-0.25[d]	0.44
	ME360$_{Samsung}$	-0.02 (-0.03, -0.01)[c]	0.07	0.11	-0.35[c]	0.58	-0.06 (-0.07, -0.04)[b]	0.06	0.12	-0.26[d]	0.40
	myBVI	-0.05 (-0.06, -0.03)[b]	0.06	0.08	0.07	0.60	-0.03 (-0.04, -0.02)[b]	0.04	0.09	-0.10	0.62
W:Height	ME360$_{Apple}$	-0.00 (-0.01, 0.01)	0.03	0.05	-0.17	0.95	-0.02 (-0.03, -0.01)[b]	0.04	0.08	-0.20[c]	0.79
	ME360$_{Samsung}$	0.00 (-0.01, 0.01)	0.03	0.06	-0.20[b]	0.94	-0.02 (-0.03, -0.01)[c]	0.04	0.08	-0.20[c]	0.78
	myBVI	-0.00 (-0.01, 0.01)	0.02	0.04	-0.02[b]	0.97	-0.02 (-0.03, -0.02)[b]	0.04	0.06	0.01	0.88

95%CI: 95% confidence interval; β: regression coefficient produced from linear regression as used to assess proportional bias; EQ: equivalence; HC: hip circumference; LOA: limits of agreement; MD: mean difference; R^2: r-squared; RMSE: root mean square error; SEE: standard error of the estimate; WC: waist circumference; WHR: waist-to-hip ratio; W:HT: waist-to-height ratio. [a] Expressed as mean (95% confidence interval); [b] Indicates statistical significance at p < 0.001; [c] indicates statistical significance at p < 0.010; [d] indicates statistical significance at p <0.05

Table 5. Agreement between smartphone-based automated anthropometrics and reference tape measurements between White and Black/African American participants

N=104		White (n=76)					Black/African American (n=28)				
		MD (95%CI)[a]	RMSE	LOA	β	R^2	MD (95%CI)	RMSE	LOA	β	R^2
WC	ME360$_{Apple}$	-1.8 (-3.2, -0.5)[c]	6.1	11.5	-0.14[d]	0.77	-3.5 (-6.1, -0.9)[d]	7.4	13.1	-0.18[c]	0.93
	ME360$_{Samsung}$	-1.0 (-2.4, 0.4)	6.3	12.2	-0.15[d]	0.75	-2.5 (-5.2, 0.1)	7.2	13.4	-0.18[c]	0.92
	myBVI	-2.7 (-3.7, -1.6)[b]	5.3	9.0	0.01	0.87	-1.6 (-3.4, 0.2)	4.9	9.2	0.02	0.96
HC	ME360$_{Apple}$	2.9 (2.2, 3.5)[b]	4.1	5.9	-0.11[c]	0.90	4.2 (2.4, 6.0)[b]	6.2	9.1	-0.09	0.93
	ME360$_{Samsung}$	3.3 (2.6, 3.9)[b]	4.3	5.5	-0.11[c]	0.92	5.6 (3.8, 7.3)[b]	7.0	8.7	-0.09	0.93
	myBVI	1.3 (0.2, 2.4)[d]	4.9	9.2	0.18[b]	0.83	4.8 (2.5, 7.2)[b]	7.7	11.9	0.14[d]	0.91
WHR	ME360$_{Apple}$	-0.04 (-0.05, -0.03)[b]	0.07	0.10	0.00	0.50	-0.06 (-0.08, -0.04)[b]	0.08	0.11	-0.19	0.55
	ME360$_{Samsung}$	-0.04 (-0.05, -0.02)[b]	0.07	0.11	-0.00	0.48	-0.06 (-0.08, -0.04)[b]	0.08	0.10	-0.26	0.58
	myBVI	-0.04 (-0.05, -0.03)[b]	0.06	0.09	0.06	0.64	-0.05 (-0.07, -0.04)[b]	0.06	0.07	0.01	0.82
W:HT	ME360$_{Apple}$	-0.01 (-0.02, -0.00)[d]	0.04	0.07	-0.15[d]	0.78	-0.02 (-0.04, -0.01)[c]	0.04	0.08	-0.20[c]	0.94
	ME360$_{Samsung}$	-0.00 (-0.01, 0.00)	0.04	0.08	-0.19[c]	0.75	-0.02 (-0.03, -0.00)	0.04	0.07	-0.20[c]	0.92
	myBVI	-0.01 (-0.02, -0.00)[b]	0.03	0.06	-0.03	0.86	-0.01 (-0.02, -0.00)	0.03	0.05	0.01	0.95

95%CI: 95% confidence interval; β: regression coefficient produced from linear regression as used to assess proportional bias; EQ: equivalence; HC: hip circumference; LOA: limits of agreement; MD: mean difference; R^2: r-squared; RMSE: root mean square error; SEE: standard error of the estimate; WC: waist circumference; WHR: waist-to-hip ratio; W:HT: waist-to-height ratio. [a] Expressed as mean (95% confidence interval); [b] Indicates statistical significance at $p < 0.001$; [c] indicates statistical significance at $p < 0.010$; [d] indicates statistical significance at $p < 0.05$
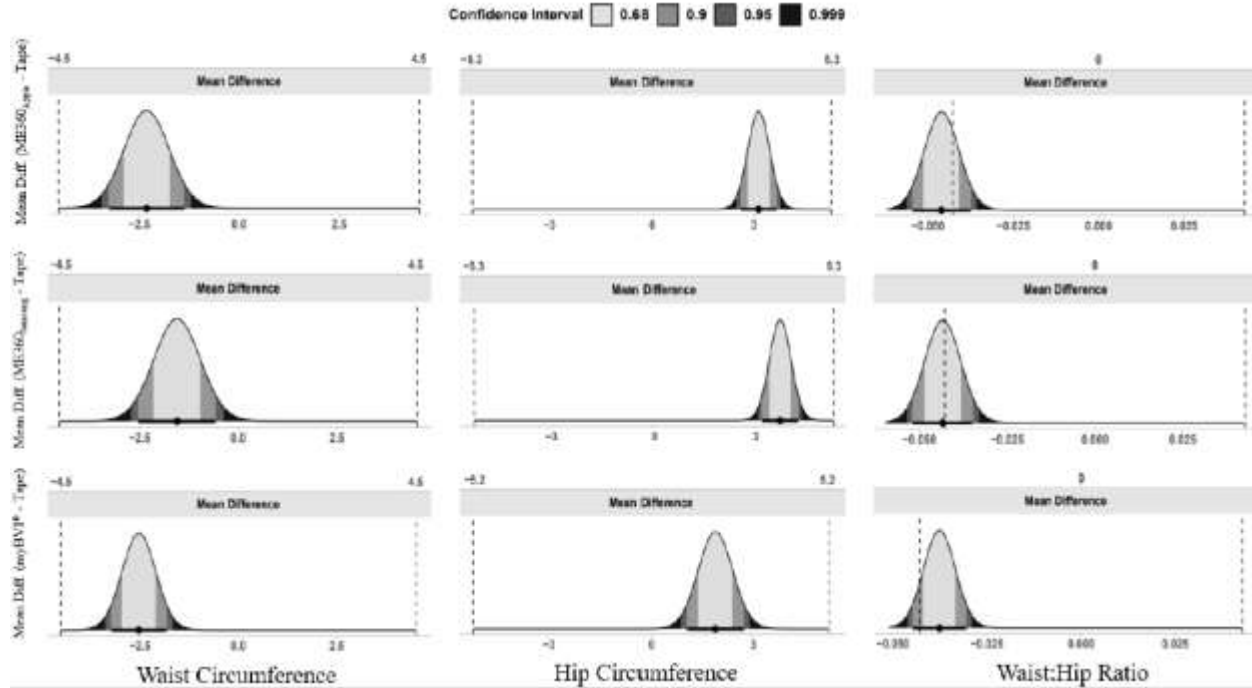
Figure 1. Equivalence Between Smartphone-based Automated Anthropometrics and Reference Tape Measurements. An illustration of the equivalence between WC, HC, and WHR produced by each mobile application and those produced by the reference is shown. Variables are considered equivalent with the tape measurements when the entire 90%CI is within the equivalence region.
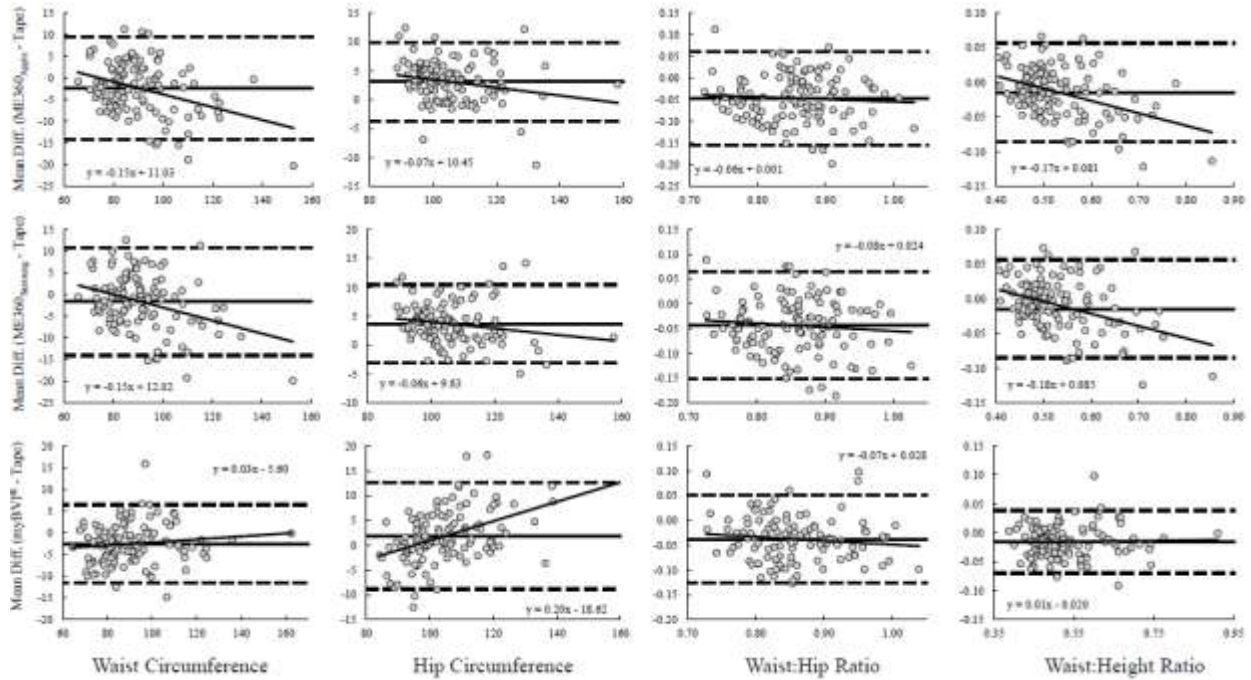
Figure 2. Bland-Altman Plots of Smartphone-based Automated Anthropometrics. Bland-Altman plots are presented. Solid diagonal line: relationship between the mean difference in circumference estimates (y-axis) and the average of the automated and tape measurements (x-axis). Solid horizontal line: average mean difference. Dashed horizontal lines: 95% limits of agreement.