Distribution Agreement

Signature:

______________________________ ______________

Joel Conkle Date

Development and Evaluation of a 3D Imaging System for Child Anthropometry

By

Joel Conkle
Doctor of Philosophy

Nutrition and Health Sciences

Reynaldo Martorell, PhD
Advisor

Eugene Alexander, PhD
Committee Member

Rafael Flores-Ayala, DrPH MApStat
Committee Member

Usha Ramakrishnan, PhD
Committee Member

Parminder Suchdev, MD MPH
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Development and Evaluation of a 3D Imaging System for Child Anthropometry

By

Joel Conkle
MPH, Tulane University, 2007
BA, University of South Carolina, 2002

Advisor: Reynaldo Martorell, PhD

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Nutrition and Health Sciences
2017

The usefulness of anthropometry is undermined by poor measurement quality, which has led to calls from the global nutrition community for new technology to improve the quality of child anthropometry. In response, a full-body 3D imaging system, AutoAnthro, was designed to measure child stature, arm circumference (MUAC) and head circumference (HC). The research that makes up this dissertation came from the Body Imaging for Nutritional Assessment Study (BINA), which was a large-scale validation study of AutoAnthro.

BINA collected manual and 3D scan-derived measurements from 474 children under five years of age in Atlanta, USA. We first analyzed manual measurement quality to confirm that we collected gold standard anthropometry. We then evaluated the reliability and accuracy of 3D scan-derived measurements against manual measurements, and included an assessment of how similar the two methods were in classifying nutritional status. Finally, we evaluated the efficiency, invasiveness, and user experience of 3D imaging by conducting a time-motion study on a subsample of BINA participants, and by interviewing BINA anthropometrists. To place our research into context we carried out literature reviews on anthropometric data quality and the use of 3D imaging for anthropometry.

After finding excellent quality of manual measurements we concluded that BINA could provide a meaningful evaluation of 3D imaging for child anthropometry. In comparing the two methods we found that measurement reliability of repeated scans was excellent, and similar to manual measurement reliability for stature, HC and MUAC. We found systematic bias when analyzing accuracy — 3D imaging overestimated stature and HC and underestimated MUAC. After adjusting scan-derived measurements to remove systematic bias, 3D imaging and manual measurement yielded similar mean z-scores, z-score standard deviations (SD), and prevalence. Sensitivity and specificity of adjusted, scan-derived measurements was good to excellent for all measures. Qualitative data showed anthropometrists considered the use of AutoAnthro an easy, 'streamlined experience' when measuring cooperative children, but scanning uncooperative children was difficult. We found that scanning took less time and was less stressful for children than manual measurement.

Technology could be the most efficient driver of anthropometric data quality improvement. We do not yet know if AutoAnthro will lead to improved quality of child measurements, but BINA showed that a 3D imaging system produced reliable measurements of children under five years of age, which suggests that 3D imaging can be an appropriate anthropometric tool for infants and young children. Further research and development is needed, particularly to determine if AutoAnthro improves quality and to address our findings of systematic inaccuracy and anthropometrists' lack of confidence in scanning uncooperative children. The potential value of 3D imaging for anthropometry is not limited to quality improvement; adoption of the technology could result in collection of hundreds of measurements during regular nutritional assessment, and lead to the discovery of new indicators that make anthropometry a better predictor of outcomes of interest.

Development and Evaluation of a 3D Imaging System for Child Anthropometry

By

Joel Conkle
MPH, Tulane University, 2007
BA, University of South Carolina, 2002

Advisor: Reynaldo Martorell, PhD

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Nutrition and Health Sciences
2017

Acknowledgements

Hundreds of parents and caregivers in the Atlanta area agreed to let their children participate in the BINA study. This research would not have been possible without their support and their children's participation. We are also grateful for the support of the staff of study site facilities.

Jennifer Breiman, RN; Carma Graham, MS; Ashton Hughes, BA; Kate Keirsey, MS; and Terrell Williams, MPH were research specialists for the study who helped to develop the study manual and tools, collected high quality anthropometry and 3D scans, participated in chapter six research as collaborators, and provided feedback for successful completion of the study and further improvement of the 3D imaging system. Their dedication was essential for the BINA study and will continue to make an impact as 3D imaging for child anthropometry is further developed.

Kenneth H. Brown, MD, The Bill and Melinda Gates Foundation, championed development of the 3D imaging system and assisted in the study design and interpretation. Dr. Brown responded to the call to improve anthropometric data quality with new technology, and along with Body Surface Translations, Inc. he made this research possible. I received funding for graduate study from the National Institutes of Health Institutional National Research Service Award, the Emory University Laney Graduate School Fellowship, and the Emory's Center of Excellence in Maternal and Child Health (MCH) Education, Science and Practice.

The dissertation advisor and committee members, listed on the approval sheet, were BINA study investigators; they spent many hours reviewing and revising this research, and they helped me develop the expertise necessary to complete the dissertation. Ideas and support from committee members vastly improved the quality of this research. I am grateful for the administrative support provided by Sandra C. Smith and Joan Lynfatt for the BINA study and for my graduate studies respectively. I am also grateful for the input of reviewers from the US Centers for Disease Control and Prevention who edited two of the papers included in the dissertation, and for the guidance of Emory professors who helped me to develop research skills on other research projects, including Matthew C. Freeman, Melissa Young, and Aryeh D. Stein.

Contents

List of Figures

List of Tables

Research in Technology Context

Evidence before this research

3D imaging systems for anthropometry have been around since the 1950s. In the 1980s technology advanced from photographs to using lasers for body digitization. By the 2000s the falling cost of 3D imaging systems (from USD $100s of thousands to less than $10,000) made commercial application feasible, and the technology was deemed accurate and reliable enough to be used in national sizing surveys of adults and older children around the globe. Multiple studies in the health sector found that the technology used in sizing surveys could reliably produce measurements for the assessment of nutritional status. In 2010 a lower cost (<USD $1,000), portable imaging technology, light-coding, was introduced to the market in the Microsoft Kinect. A few studies evaluated the use of the Kinect for adult anthropometry and found good reliability. The vast majority of studies and research using 3D imaging for anthropometry did not include children under four years of age because imaging systems were not designed to handle movement. Prior to our research, there was one 3D imaging system designed for young children, StarScanner, which used technology similar to that used in sizing surveys, and was designed specifically to scan a newborn's head to design orthoses. 3D imaging is not used in the health sector for regular nutritional assessment of adults or children.

Added value of this research

In 2013 light-coding technology was placed in an open-source, handheld scanner; the Structure Sensor. In this research we evaluated the AutoAnthro 3D imaging system, which uses a Structure Sensor scanner and custom software. To our knowledge AutoAnthro is the first 3D imaging system designed specifically for full-body anthropometry of infants and young children. Our research is the first evaluation of AutoAnthro, and to our knowledge it is the first research on using a single, handheld scanner for anthropometry. AutoAnthro and StarScanner share the same capture strategy for handling movement — taking multiple scans of short duration and stitching them together. Our research showed that the capture strategy worked well for newborns, infants under six months of age, and children three years of age and over; but that more software development is needed to make anthropometrists comfortable in scanning children six months to three years of age who are not cooperative. Our findings on the reliability and accuracy of scan-derived anthropometry of children under five were similar to previous studies in the health sector on adults, and those studies concluded that 3D imaging could be used as a measurement tool. We also found that 3D imaging was less stressful than manual measurements for children, which was not assessed in the literature that we reviewed.

Implications of all the available evidence

3D imaging is not a new idea for anthropometry, but there have been significant barriers to regular use of the technology in the health sector. The high cost and requirement for dedicated space to setup multiple cameras in fixed locations meant that 3D imaging was not appropriate for regular nutritional assessment and screening. For children, an additional barrier was that 3D imaging systems were not designed for child anthropometry. The development of a low-cost, handheld, single, scanner addressed barriers and opened the door to making 3D imaging a common tool for adult anthropometric data collection. The development of AutoAnthro and our research showed that 3D imaging also has the potential to be a useful tool for nutritional assessment and screening of children. Based on findings from a single study we cannot yet make policy or practice recommendations. More research on portable 3D imaging systems, including AutoAnthro, is needed before we can make recommendations on replacing manual measurement equipment with 3D imaging for infants and young children.

Chapter 1 . Introduction

Anthropometry, or measurement of the human body, is an ancient practice. Texts from Ayurvedic and Traditional Chinese Medicine show that human beings have attributed meaning to variation in human surface morphology for thousands of years (2, 3). Anthropometric methods were standardized in science in the 18th and 19th centuries, and while interpretation of anthropometry changed dramatically over the centuries, with measurements applied to health and wellbeing, productivity and fighting ability, fortune-telling, and eugenics; methods for anthropometry changed little since the 1800s (3, 4). Today's anthropometric tools are rudimentary. We rely on wooden boards, tapes and calipers; which are some of the same basic tools found in B.C. China (3). These basic tools are used to measure body length, height and various circumferences and skinfold thicknesses; and anthropometric data are used to calculate indices, such as height-for-age and body mass index, which are compared to a reference population and associated to disease risk to give meaning to the measurements. Anthropometric indicators provide a direct assessment of body size, and are also routinely used to indirectly estimate body composition (fat, muscle, etc.). However, anthropometric indicators are not good proxies for body composition. Fortunately, modern nutritional assessment is not limited to anthropometry, and there have been numerous advances in measuring body composition. We now have the ability to more directly measure body composition through laboratory techniques such as air displacement plethysmography, stable isotope dilution, dual energy x-ray absorptiometry, and neutron activation analysis

(5). New technology improved measurement of body composition, but did little to improve anthropometry.

1.1 Calls for Improved Anthropometric Data Quality

The lack of advancement in anthropometric methods is problematic. First, we still rely on anthropometry as a poor proxy for body composition because the cost and complexity of laboratory techniques make them unsuitable for routine assessment. Second, current equipment, especially length/height boards, are bulky and heavy, which places a burden on anthropometrists in remote areas who are asked to carry this equipment from village to village and house to house. Third, and most importantly for our research, current methods are susceptible to human error (6) and when applied outside of a research setting often result in poor quality anthropometric data.

Skinfold thicknesses and some circumference measurements are particularly unreliable (7-9), and accurate measurement of child length is often a problem (1). Poor quality child anthropometry is common in both health facilities and surveys, and the issue is not limited to low-resource settings (9-13). Anthropometric data quality was evaluated extensively in large-scale surveys in developing countries, such as the Demographic and Health Survey (DHS) and the Multiple Indicator Cluster Survey (MICS), which routinely measure weight and length/height. Evaluations covering hundreds of surveys in developing countries found too many biologically implausible measurements (11-13) and overdispersion of length/height-for-age z-scores (HAZ) (11). Overdispersion, or too much variability, is a result of poor reliability, and causes overestimation of prevalence. A recent study found that DHS and MICS carried out in Western and Central Africa from 1990-2012 may have overestimated the prevalence of stunting (HAZ <-2 SD) by ~10 percentage points (14). Poor quality was attributed to incorrect age and inaccurate measurement of stature, particularly length (11). Stature is prone to measurer error because current methods require correct positioning of the child and correct reading of the measurement. Measurer error is exacerbated by an uncooperative child — it can be difficult to get young children to stay still during measurement, and children under two often actively resist the stature measurement because they are distressed by the requirement to lie down on a length board.

Anthropometric data quality varies between countries and between surveys in the same country; making it difficult to meaningfully compare countries, analyze trends over time, or target public health interventions. At the individual level poor quality limits our ability to monitor growth and leads to misclassification of nutritional status. The usefulness of anthropometry is undermined by poor measurement quality, which has led to calls from the global nutrition community for new technology to improve the quality of child anthropometry (11, 15).

1.2 Overview of Research Purpose, Materials and Methods

The Bill and Melinda Gates Foundation responded to the call for improved anthropometric methods by supporting Body Surface Translations Inc. (BST) to develop AutoAnthro, a full-body 3D imaging system designed to measure child stature, arm circumference (MUAC) and head circumference (HC). BST had previous experience estimating accurately and the weight of pigs using surface morphology derived from 3D imaging as a predictor; weight is a key determinant of readiness for the market and 3D imaging made this measurement easier. BST partnered with nutrition experts at Emory University to assist in technology development and to evaluate AutoAnthro. The research that makes up this dissertation came from the Body Imaging for Nutritional Assessment Study (BINA), which was a large-scale validation study of AutoAnthro. BINA collected manual and 3D scan-derived measurements from 474 children under five years of age in Atlanta, USA.

In order to draw conclusions on the ability of 3D imaging to accurately measure children, BINA needed to collect gold-standard manual anthropometry. Our first paper, chapter three of the dissertation, determined if the quality of manual anthropometry collected in BINA was good enough to be considered gold-standard. We evaluated manual anthropometry quality by examining digit preference, biological plausibility of z-scores, z-score standard deviations, and reliability; comparing results to the quality of manual anthropometry used to develop the 2006 WHO Child Growth Standards, and to various standards of anthropometric data quality. In our first paper we also provided detailed methodology for manual anthropometry data collection in BINA, and gave recommendations on how to collect high quality manual anthropometry and how to improve the assessment of anthropometric data quality.

In chapters four and five, our second paper, we answered BINA's core research question on whether or not AutoAnthro provided accurate and reliable measurements of children under five years of age. We evaluated the reliability and accuracy of 3D scan-derived measurements against manual measurements, and included an assessment of how similar the two methods were in classifying nutritional status. In chapters four and five we also included detailed methodology on producing measurements with 3D scanning, along with recommendations for further research.

In our third and final paper, chapter six, we evaluated the efficiency, invasiveness, and user experience of AutoAnthro. For our third paper we conducted a time-motion study on a subsample of BINA participants, and qualitative interviews of BINA anthropometrists; adopting a mixed-methods (quantitative and qualitative), collaborative (designed and implemented with anthropometrists) approach to provide a comprehensive assessment. We compared measurement time for scanning against manual measurement, and analyzed interviews using grounded theory. Paper 3 also provided recommendations for improvements to AutoAnthro.

To place our research in context we carried out literature reviews on anthropometric data quality and the use of 3D imaging for anthropometry. For the latter we searched PubMed for studies including "3D" and "anthropometry" in the titles, which returned 13 results. We reviewed all papers without restriction, including date and language. We used references from reviewed papers and expert advice to identify 33 additional sources. By using PubMed we intentionally biased our literature review towards the use of 3D imaging in the health sector. For the literature review on anthropometric data quality we primarily relied on references from known papers and expert advice. We also searched PubMed for "anthropometry" and "accuracy" or "reliability."

Chapter 2 . Background

2.1. 3D Imaging for Anthropometry – The Past (1800s and 1900s)

Photogrammetry can be traced back to the 1800s when Laussedat used photographs for topographic mapping (16). In the 1940s and 1950s anthropologists interested in somatotyping developed and tested standardized procedures of photogrammetry for anthropometry (17). Early photogrammetry used one camera, deriving measurements from a 2D image based on a background that provided scale (17). It was quickly recognized that there were limitations to measuring a 3D human from a 2D image, and as early as 1952 methods were developed to use a pair of facial photographs to create a rough, 3D representation of a human face (18). The use of 2D photographs for 3D reconstruction; referred to as stereo photogrammetry, 3D photogrammetry, or biostereometrics when applied to biology; required the use of multiple cameras and a calibration object (18). Manual identification of landmarks or grid points in early stereo photogrammetry was labor intensive, but processing was eventually automated (19). One of the earliest uses of automated processing of 3D photogrammetry for anthropometry was to assess the nutritional status of astronauts; the researcher used a Cray supercomputer and software from the United States Air Force (USAF) that was designed for aerial mapping (M. Golden, personal communication, April 18, 2017, (20)). In the late 1960s Moire pattern imaging, a photographic method utilizing projection of grids, was used for topographic maps;

and by the late 1970s the technique was applied to anthropometry in research on plastic and reconstructive surgery (19).

In the 1980s 3D imaging for anthropometry advanced from photographs to body digitization with lasers via ‘range imaging’, which is a blanket term covering various methods that project light onto the person being measured and use triangulation to construct a 3D surface map (19, 21). A number of terms are used to describe range imaging methods, including laser scanning and structured light (19). Like with the 3D photogrammetry, the USAF was again involved in technology development for range imaging. The USAF worked with Anthropology Research Project Inc. and the Idaho National Engineering Laboratory to adapt commercial scanning technology to anthropometry, and in 1986 the USAF helped Cyberware Inc. (dissolved in 2011) adapt their range imaging system (designed to make 3D sculptures) to anthropometry (22). Around the same time in the UK another range imaging system was developed, the Loughborough Anthropometric Shadow Scanner, and like in the US the system was developed through a public private collaboration, but in the UK the collaboration involved the garment manufacturing industry (23). In both the US and the UK the major driver for the development of a 3D imaging system for anthropometry was that sizing surveys, needed for garment design and ergonomics, were labor intensive (22, 23). The garment industry wanted sizing surveys with large samples (4500 to 6500) and 40 measurements per subject — with these requirements manual methods took too much time and money (23). By the late 1990s a large scale sizing survey, the Civilian American and European Surface Anthropometry Resource (CAESAR), employed the use of 3D imaging for anthropometry for a target sample size of ~12,000 (22). The CAESAR survey collected data in the US, the Netherlands, and Italy using a Cyberware scanner in the US and a Vitronic (Wiesbaden, Germany) scanner in Europe. The survey was supported by more than 20 industrial partners, primarily garment and automotive, and was overseen by the Air Force Research Laboratory with technical support from the Netherlands and Loughborough University (22).

2.2. 3D Imaging for Anthropometry – 2000-Present

By the 2000s the cost of 3D range imaging systems dropped dramatically, from USD $100s of thousands to under ten thousand dollars; and 3D imaging was common in national sizing surveys around the globe (21, 24-26). In 2001 SizeUK used the TC2 scanner ([TC]², Cary, USA), and multiple countries then used the same or similar technology and the same naming convention for their own survey, giving us SizeUSA, SizeJapan, SizeKorea, SIzeThailand, and others (24, 25, 27). Researchers started to use the 3D data from sizing surveys for health applications. In 2007 Wells et al examined associations between body shape and body mass index using SizeUK data (24), and researchers in Thailand used SizeThailand data to study diabetes and obesity (27).

Although there is more advanced imaging available in the health sector, such as computed tomography and magnetic resonance imaging, interest remains in 3D imaging systems because of lower cost and no radiation exposure (19). Researchers used photograph or scan-derived anthropometry to diagnose scoliosis, underdevelopment of the optic nerve, and melanoma; to assess treatment of skin ulcers, and to predict obstructive sleep apnea (28). Researchers have also tested commercial range imaging scanners for measurements relevant to the assessment of nutritional status, such as height (29), circumferences (30-32), body surface and volume (24, 33, 34), and body shape (24, 35, 36). Some of the studies considered the use of 3D imaging in nutritional epidemiology. Jaeschke et al found that scan-derived measurements correlated as well as manual measurements with biochemical markers of metabolic syndrome (31), and Lin et al developed a new index from scan-derived waist, breast and hip area; the Health Index; and found good correlation between the new index and biochemical markers of metabolic disorders (36). Scanner technology used in studies on assessment of nutritional status were the same or similar to those used in sizing surveys. Sizing surveys and research over the last decade showed that 3D scanners are a promising tool for anthropometry, but the use of 3D scanners for nutritional assessment is not common outside of research in the health sector. Commercial scanners are used in anthropology and forensics (37), and in the health sector for orthotics and orthodontics (38, 39), but we do not yet find 3D scanners used in primary healthcare clinics or health and nutrition surveys.

The experience of the Body Benchmark Study provides insight into why the 3D imaging systems that are commonly used for sizing surveys are not yet common in the health sector. In the early 2000s Select Research (Worcestershire, England), a company that carried out sizing surveys for the garment industry, researched applications for 3D imaging in the health sector (40). In 2007, in collaboration with the Mayo Clinic and Heartlands Hospital, Select Research launched the Body

Benchmark Study, a study that set out to replace anthropometric proxies of body composition, specifically BMI, with a new 3D measurement derived from 3D scans, the Body Volume Index (BVI) (40). The study results were made public in 2010 and the overall conclusion was that BVI offered advantages over traditional measures (40). In 2010 the NHS reviewed the research and rejected a proposal to install 3D scanners across the NHS; the company reported that they were advised by the NHS to develop a low-cost, mobile solution (40). The imaging systems used by Select Research, sizing surveys, and the medical research described in the previous paragraph require multiple cameras in a fixed position for stereo triangulation, and a long-scanning period (~10 seconds); and while the cost of such a system may be reasonable for some uses, it was too expensive for implementation across primary healthcare centers.

In the 1990s in a review of digital photogrammetry Mitchell concluded that the requirements for the use of photogrammetry in the health sector are a "surprisingly low level of cost" and no requirement for a "work-station," pointing out that surface morphology is not crucial to a patient's health because of the availability of internal examination (41). The same barriers applied to the attempt to bring range imaging technology to the NHS in 2010. An additional limitation of range imaging systems was that the long scanning period made it difficult to measure children, especially children under three years of age that move constantly. In the 1950s Dupertuis, a man interested in standardizing anthropometric photogrammetry for somatotyping, reported that the "technique seems admirably suited to longitudinal growth study after the first few years (17)." Dupertuis limited his tests to adults and children over three years of age. By 2010 3D imaging was not yet used for monitoring child growth and there were no solutions for full body imaging of infants and young children.

In this decade a 3D scanner was specifically designed for children. StarScanner (Vorum, Vancouver, BC, Canada) is an US FDA approved medical device to scan a newborn's head for cranial remolding orthoses (39); the device was licensed to Orthomerica and is now in hundreds of medical facilities. The StarScanner, like the systems before it, requires fixed, multiple cameras. It was not a low-cost, mobile solution, but it made its way into the health sector by fitting a specialized need. In 2010 PrimeSense (acquired by Apple in 2013) licensed its "light-coding" technology for use in the Microsoft Kinect. Light-coding is range imaging that requires a single device: an infrared projector and sensor are contained in the same device and stereo triangulation is achieved by comparing the sensor image to an image of the projector's pattern that is hardwired into a microchip. Light-coding reduced the cost and size of 3D scanners and led to the use of 3D imaging in the gaming industry. A few studies were carried out to evaluate the use of Microsoft Kinect for anthropometry. One study measured stationary cylinders as a proxy for human circumferences (43); another study compared Kinect to a more expensive range imaging system for various measurements (16); and a third study made estimates of body volume with the Kinect (16, 44). None of the studies included infants or children under five years of age, and they all used multiple Kinect devices in fixed positions. In 2013 a Kickstarter campaign funded the development of

Structure Sensor (Occipital, San Francisco, CA, USA), an open-source, light-coding 3D scanner that attaches to a tablet or phone (Figure 2-1).

A. 

B. 

Figure 2-1 A. Hardware setup for the Body Imaging for Nutritional Assessment Study — Structure Sensor connected to tablet. B. Structure Sensor connected to mobile phone

The development of low-cost, ultra-portable, open-source scanners provided an opportunity to extend the use of 3D imaging to common uses of anthropometry in the health sector, such as nutritional screening and surveillance. The open-source nature of Structure Sensor makes it ideal for the development of new software. AutoAnthro, the 3D imaging system evaluated in this dissertation, is made up of custom software from BST for the Structure Sensor scanner. Previously, BST worked with a large, custom 3D scanner that was carried on the back of the operator, and their imaging system required that the pigs be snared to prevent movement. AutoAnthro uses a single, handheld scanner that is not in a fixed position, and the system was developed to allow for some movement with regular nutritional assessment and growth monitoring of newborns, infants and young children in mind.

References (Chapters 1 and 2)

1. Lipman TH, Hench KD, Benyi T, Delaune J, Gilluly KA, Johnson L, et al. A multicentre randomised controlled trial of an intervention to improve the accuracy of linear growth measurement. Arch Dis Child. 2004;89(4):342-6.

2. Thakur V, Bhatnagar V. ANTHROPOMETRY IN AYURVEDA W.S.R. TO SUŚRUTOKTA AṄGULA PRAMĀṆA ŚĀRĪRA2017. 440-3 p.

3. Shi B, Wu, C. Anthropometry in Ancient China--Based on the Physiognomy Book of Divine Fortuneteller Ma Yi. 2nd Annual International Conference on Social Science and Contemporary Humanity Development; Wuhan, Hubei, China: Atlantis Press; 2016.

4. Ulijaszek S, Komlos J. From a History of Anthropometry to Anthropometric History. Human Variation. Society for the Study of Human Biology: CRC Press; 2010. p. 183-97.

5. Heymsfield SB, Ebbeling CB, Zheng J, Pietrobelli A, Strauss BJ, Silva AM, et al. Multi-component molecular-level body composition reference methods: evolving concepts and future directions. Obesity Reviews. 2015;16(4):282-94.

6. WHO Multicentre Growth Reference Study Group. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl. 2006;450:38-46.

7. Sebo P, Haller D, Pechere-Bertschi A, Bovier P, Herrmann F. Accuracy of doctors' anthropometric measurements in general practice. Swiss Med Wkly. 2015;145:w14115.

8. Sebo P, Beer-Borst S, Haller DM, Bovier PA. Reliability of doctors' anthropometric measurements to detect obesity. Prev Med. 2008;47(4):389-93.

9. Yin H, Dai Y, Li H, Xie X, Ren H. The test–re-test reliability of routine infant anthropometry at primary care hospitals in Chongqing, PR China. Annals of Human Biology. 2013;40(4):309-17.

10. Gerner B, McCallum Z, Sheehan J, Harris C, Wake M. Are general practitioners equipped to detect child overweight/obesity? Survey and audit. J Paediatr Child Health. 2006;42(4):206-11.

11. Assaf S, Kothari, M., Pullum, T. An Assessment of the Quality of DHS Anthropometric Data, 2005-2014. DHS Methodological Reports No 16. Rockville, MD, USA: ICF International; 2015.

12. Pullum TW. An Assessment of the Quality of Data on Health and Nutrition in the DHS Surveys, 1993-2003. Methodological Reports No 6. Calverton, MD, USA: Macro International Inc.; 2008.

13. Corsi DJ, Perkins JM, Subramanian SV. Child anthropometry data quality from Demographic and Health Surveys, Multiple Indicator Cluster Surveys, and National Nutrition Surveys in the West Central Africa region: are we comparing apples and oranges? Glob Health Action. 2017;10(1):1328185.

14. Grellety E, Golden MH. The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition. PloS one. 2016;11(12):e0168585.

15. USAID. Anthropometric Data in Population-Based Surveys, Meeting Report, July 14-15, 2015. Washington, D.C., USA: FHI 360/FANTA; 2016.

16. Soileau L, Bautista D, Johnson C, Gao C, Zhang K, Li X, et al. Automated anthropometric phenotyping with novel Kinect-based three-dimensional imaging method: comparison with a reference laser imaging system. Eur J Clin Nutr. 2016;70(4):475-81.

17. Dupertuis CW, Tanner JM. The pose of the subject for photogrammetric anthropometry, with especial reference to somatotyping. American Journal of Physical Anthropology. 1950;8(1):27-48.

18. Douglas TS. Image processing for craniofacial landmark identification and measurement: a review of photogrammetry and cephalometry. Computerized Medical Imaging and Graphics. 2004;28(7):401-9.

19. Riphagen JM, van Neck JW, van Adrichem LN. 3D surface imaging in medicine: a review of working principles and implications for imaging the unsedated child. J Craniofac Surg. 2008;19(2):517-24.

20. Whittle MW, Herron RE, Cuzzi JR. Biostereometric analysis of body form: the second manned Skylab mission. Aviat Space Environ Med. 1976;47(4):410-2.

21. D'Apuzzo N. Recent advances in 3D full body scanning with applications to fashion and apparel. In: Gruen A, Kahmen, H., editor. Optical 3D Measurement Techniques IX; Vienna, Austria2009.

22. Robinette KM, Daanen H, Paquet E, editors. The CAESAR project: a 3-D surface anthropometry survey. Second International Conference on 3-D Digital Imaging and Modeling (Cat NoPR00062); 1999 1999.

23. Jones PR, West GM, Harris DH, Read JB. The Loughborough anthropometric shadow scanner (LASS). Endeavour. 1989;13(4):162-8.

24. Wells JC, Treleaven P, Cole TJ. BMI compared with 3-dimensional body shape: the UK National Sizing Survey. Am J Clin Nutr. 2007;85(2):419-25.

25. Kim JY, You JW, Kim MS. South Korean anthropometric data and survey methodology: 'Size Korea' project. Ergonomics. 2017:1-11.

26. Charoensiriwath S, editor A real-time data monitoring and management system for Thailand 2019 first national sizing survey. PICMET '08 - 2008 Portland International Conference on Management of Engineering & Technology; 2008 27-31 July 2008.

27. Charoensiriwath S, editor SizeThailand e-Health: A personalised health monitoring and diagnosis system using 3D body scanning technology. PICMET 2010 TECHNOLOGY MANAGEMENT FOR GLOBAL ECONOMIC GROWTH; 2010 18-22 July 2010.

28. Ey-Chmielewska H, Chrusciel-Nogalska M, Fraczak B. Photogrammetry and Its Potential Application in Medical Science on the Basis of Selected Literature. Adv Clin Exp Med. 2015;24(4):737-41.

29. Kuehnapfel A, Ahnert P, Loeffler M, Broda A, Scholz M. Reliability of 3D laser-based anthropometry and comparison with classical anthropometry. Scientific reports. 2016;6:26672.

30. Pepper MR, Freeland-Graves JH, Yu W, Stanforth PR, Cahill JM, Mahometa M, et al. Validation of a 3-dimensional laser body scanner for assessment of waist and hip circumference. Journal of the American College of Nutrition. 2010;29(3):179-88.

31. 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;10(3):e0119430.

32. Wang J, Gallagher D, Thornton JC, Yu W, Horlick M, Pi-Sunyer FX. Validation of a 3-dimensional photonic scanner for the measurement of body volumes, dimensions, and percentage body fat. Am J Clin Nutr. 2006;83(4):809-16.

33. Adler C, Steinbrecher A, Jaeschke L, Mahler A, Boschmann M, Jeran S, et al. Validity and reliability of total body volume and relative body fat mass from a 3-dimensional photonic body surface scanner. PloS one. 2017;12(7):e0180201.

34. Garlie TN, Obusek JP, Corner BD, Zambraski EJ. Comparison of body fat estimates using 3D digital laser scans, direct manual anthropometry, and DXA in men. Am J Hum Biol. 2010;22(5):695-701.

35. Loffler-Wirth H, Willscher E, Ahnert P, Wirkner K, Engel C, Loeffler M, et al. Novel Anthropometry Based on 3D-Bodyscans Applied to a Large Population Based Cohort. PloS one. 2016;11(7):e0159887.

36. Lin JD, Chiou WK, Weng HF, Fang JT, Liu TH. Application of three-dimensional body scanner: observation of prevalence of metabolic syndrome. Clin Nutr. 2004;23(6):1313-23.

37. Owaydhah WH, Alobaidy MA, Alraddadi AS, Soames RW. Three-dimensional analysis of the proximal humeral and glenoid geometry using MicroScribe 3D digitizer. Surgical and Radiologic Anatomy. 2017;39(7):767-72.

38. Taneva E, Kusnoto B, Evans CA. 3D Scanning, Imaging, and Printing in Orthodontics. In: Bourzgui F, editor. Issues in Contemporary Orthodontics. Rijeka: InTech; 2015. Ch. 09.

39. Weathers WM, Khechoyan D, Wolfswinkel EM, Mohan K, Nagy A, Bollo RJ, et al. A novel quantitative method for evaluating surgical outcomes in craniosynostosis: pilot analysis for metopic synostosis. Craniomaxillofac Trauma Reconstr. 2014;7(1):1-8.

40. Barnes R, editor The Body Volume Index (BVI): Using 3D Scanners to Measure and Predict Obesity. International Conference on 3D Body Scanning Technologies; 2010; Lugano, Switzerland: Hometrica Consulting.

41. Mitchell HL. Applications of digital photogrammetry to medical investigations. ISPRS Journal of Photogrammetry and Remote Sensing. 1995;50(3):27-36.

42. Wells JC, Stocks J, Bonner R, Raywood E, Legg S, Lee S, et al. Acceptability, Precision and Accuracy of 3D Photonic Scanning for Measurement of Body Shape in a Multi-Ethnic Sample of Children Aged 5-11 Years: The SLIC Study. PloS one. 2015;10(4):e0124193.

43. Clarkson S, Wheat J, Heller B, Choppin S. Assessment of a Microsoft Kinect-based 3D scanning system for taking body segment girth measurements: a comparison to ISAK and ISO standards. J Sports Sci. 2016;34(11):1006-14.

44. Cook T. Using the Microsoft Kinect for patient size estimation and radiation dose normalization: proof of concept and initial validation. 2013.

Chapter 3 . Improving the Quality of Child Anthropometry: Manual Anthropometry in the 3D Body Imaging for Nutritional Assessment Study (BINA)

Joel Conkle, MPH[a*], Usha Ramakrishnan, PhD[a,b], Rafael Flores-Ayala, DrPH[a,c], Parminder S. Suchdev, MD MPH[a,b,c,d], and Reynaldo Martorell, PhD[a,b]

[a] Doctoral Program in Nutrition and Health Sciences, Laney Graduate School, Emory University, Atlanta, GA, USA; joel.conkle@emory.edu

[b] Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA

[c] Division of Nutrition, Physical Activity and Obesity; National Center for Chronic Disease Prevention and Health Promotion, U.S. Centers for Disease Control and Prevention, Atlanta, GA, USA

[d] Department of Pediatrics, School of Medicine, Emory University. Atlanta, GA, USA

* Correspondence: joelconkle@gmail.com; Tel.: +1-404-725-7138

Abstract:

Anthropometric data collected in clinics and surveys are often inaccurate and unreliable. The Body Imaging for Nutritional Assessment Study (BINA) evaluated the ability of 3D imaging to correctly measure stature, head circumference (HC) and arm circumference (MUAC) for children under five years of age. This paper describes the protocol for and the quality of manual anthropometric measurements in BINA, a study conducted in 2016-17 in Atlanta, USA. Quality was evaluated by examining digit preference, biological plausibility of z-scores, z-score standard deviations, and reliability. We calculated z-scores and analyzed plausibility based on the 2006 WHO Child Growth Standards (CGS). For reliability, we calculated intra- and inter-observer Technical Error of Measurement (TEM) and Intraclass Correlation Coefficient (ICC). We found low digit preference; 99.6% of z-scores were biologically plausible, with z-score standard deviations ranging from 0.92 to 1.07. Total TEM was 0.40 for stature, 0.28 for HC, and 0.25 for MUAC in centimeters. ICC ranged from 0.99 to 1.00. The quality of manual measurements in BINA was high and similar to that of the anthropometric data used to develop the WHO CGS. We attributed high quality to vigorous training, motivated and competent field staff, reduction of non-measurement error through the use of technology, and reduction of measurement error through adequate monitoring and supervision. Our anthropometry measurement protocol, which builds on and improves upon the protocol used for the WHO CGS, can be used to improve anthropometric data quality. The discussion illustrates the need to standardize anthropometric data quality assessment, and we conclude that BINA can provide a valuable evaluation of 3D imaging for child anthropometry because there is comparison to gold-standard, manual measurements.

3.1. Introduction

The Multicenter Growth Reference Study (MGRS) and subsequent development of World Health Organization (WHO) Child Growth Standards (CGS) in 2006 provided a single set of reference measurements for children around the globe (1). The WHO CGS have been commonly adapted for routine use in low- and middle-income countries as well as in some high-income countries, including the US, for purposes of individual growth monitoring, clinical research, and public health program monitoring (2). With the new reference came a single measuring protocol that could be adopted in health facilities and surveys everywhere (3); however, the MGRS protocol requires extensive training and supervision, and repeated measurements (3). The full MGRS protocol is not routinely followed in health facilities, nor is it used in large-scale surveys that evaluate nutrition such as Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS). The non-digital, manual measurements currently in use are susceptible to human error (4), and the use of inadequate measuring protocols can increase measurement error. Poor quality child anthropometry is common in both health facilities and surveys (5-9).

There is recognition that the quality of child anthropometry should be assessed before using and disseminating the data. DHS, MICS and Standardized Monitoring and Assessment of Relief and Transitions (SMART) Survey methodologies include assessment of anthropometric data quality. The assessments are all loosely based on recommendations from a WHO expert committee convened in 1995 (10), but there are methodological differences. Recently, there have been calls to improve anthropometry quality through the use of technology, and to revisit the 1995 WHO recommendations to standardize data quality assessment (11).

The data we analyze is part of the Body Imaging for Nutritional Assessment Study (BINA), which compared a 3D imaging system to currently recommended non-digital, manual measurements of stature (length and height), arm circumference (MUAC) and head circumference (HC) in 474 children under five years of age. In order to draw conclusions on the ability of 3D imaging to replace manual anthropometry, the study needed to collect gold-standard manual anthropometry. This paper describes the training, standardization and data collection methods for manual anthropometry in BINA; evaluates the quality of BINA manual anthropometry; and provides some recommendations for achieving gold standard manual anthropometry and standardizing data quality assessment to guide clinicians, researchers and public health program managers.

3.2. Materials and Methods

Primary caregivers of all children participating in BINA gave written, informed consent and the study was approved by the Emory Institutional Review Board.

3.2.1. Field Staff Training and Standardization

All five field staff selected for the BINA Study held college degrees, with three of the five holding a master’s degree at the time of the study. In August 2016 field staff completed a three-week training led by trainers from Emory University who had extensive experience with anthropometry in clinic, survey, and research settings; including experience in the study used to develop the 2006 WHO CGS. Training consisted of theoretical and practical sessions on 3D imaging and manual measurements, and field staff were trained to function as both anthropometrists and assistants. Emory University faculty and field staff developed training materials and a study manual.

Training culminated with a three-day standardization test for manual anthropometry at a local daycare center, which consisted of a lead anthropometrist from Emory University and all field staff taking repeated measurements of ten children under five years of age. As anthropometrists, field staff were assessed on all measurements except weight, and data was analyzed using ENA Software 2011 (12) and a Microsoft Excel Spreadsheet from the United States Centers for Disease Control and Prevention (US CDC) Micronutrient Survey Toolkit (13). Passing or failing the standardization test was based on accuracy and reliability results for the main measurements of interest (length and height). We determined accuracy by comparing each anthropometrist to the lead anthropometrist and to the mean of all anthropometrists. For reliability, we computed the Technical Error of Measurement (TEM), which determines if anthropometrists get similar results when carrying out repeated measurements. We also used the US CDC spreadsheets to visually present accuracy and reliability results to the anthropometrists, and to assess whether or not an anthropometrist's first measurement was systematically lower or higher than the second measurement, also known as the measurement effect. There was no evidence of substantial measurement effect. The accuracy and reliability of the lead anthropometrist and all anthropometrists for length and height were similar to results from the MGRS (3) and were classified as the highest ranking of "good" (<0.4 cm) according to the SMART suggested cut-off point of <0.4 cm for both intra-observer TEM and bias from an expert (14). Based on results of the standardization test, we retained all anthropometrists for the study.

3.2.2. Sampling, Measurement and Data Entry Procedures

Utilizing convenience sampling we recruited and measured children in facilities, primarily daycare centers, in Atlanta, GA, USA from September 2016 to February 2017. Detailed methods for participant selection and sample characteristics will be published alongside findings on 3D imaging accuracy and reliability.

For manual anthropometry we followed measurement techniques used to develop the 2006 WHO Growth Standards (3), including measuring children under two years of age lying down and measuring to the nearest tenth of a kg and cm. We measured weight with digital scales with taring function (Rice Lake Weighing Systems, Inc., Rice Lake, WI), stature with an infant/child/adult wooden board (Weigh and Measure, LLC, Maryland USA), and circumferences with synthetic measuring tapes. We routinely checked calibration of scales using known weights and replaced damaged measuring tapes. The BINA study manual includes detailed descriptions of techniques and equipment used in the study and is available upon request.

During data collection sessions, the five field staff split into two teams of two, with the fifth person acting as a floater. The techniques for scans and manual measurements required an assistant, and teammates alternated as the anthropometrist and assistant. Anthropometrists also took turns as a floater, and the main roles of the floater were to prepare children for measurement and to act as a second assistant for younger children. Anthropometry procedures designed for household surveys often contain a role for the child's caretaker. However, since we obtained data in settings where the caretakers were not present, a second assistant was needed to hold and/or position the child.

After taking time to acclimate the child and establish rapport, and after undressing the child (to their diaper or to skin-tight shorts/leotards provided by the study), the field team started with 3D scans which was then followed by manual measurements. The first anthropometrist completed one session, measuring head circumference, MUAC, and length or height. Instead of one anthropometrist completing two sessions concurrently, the field team reversed roles and a new anthropometrist completed her first session of manual measurements. After both team members completed their first session, the process was repeated for the second session. We staggered manual measurements to reduce the likelihood of anthropometrists remembering their first measurement. To further minimize bias in inter-observer error, anthropometrists entered their own measurement results and did not inform the assistant of the number, which is different from the typical anthropometry procedure of the assistant recording results on behalf of the anthropometrist.

3.2.3. Quality Control and Data Cleaning

Our custom software for electronic data capture included range checks for non-digital measurements to catch data entry errors. If a measurement was below the 0.01 percentile or above the 99.9 percentile, a pop-up box appeared on the screen and the anthropometrist could either re-enter or accept the value. The software also included automatic triggering of a third measurement for non-digital measurements. The triggers were programmed based on the MGRS standards for the maximum allowable difference (≤0.5 cm for head and arm circumference, and ≤0.7 cm for length) (3), but our study differed from MGRS in that we triggered within observer differences while MGRS triggered between observers. For each child, two anthropometrists entered demographic information, including date of birth, separately. Double data entry allowed us to identify discrepancies caused by data entry error, and we referred to the original consent form to make corrections.

For manual anthropometry, field staff received ongoing monitoring and supervision. This included examining digit preference, the percent of intra-observer measurements exceeding the maximum allowable difference, measures of inter-observer reliability namely TEM, and estimates of bias by comparing to an expert anthropometrist who measured a subsample of children. In addition to comparing to an expert anthropometrist, field staff also periodically compared with each other. Field staff regularly received data quality reports and supervision from expert anthropometrists.

3.2.4 Quality Tests

For digit preference, we examined the proportion of non-digital, manual measurements (height, head circumference and arm circumference) with each of the possible digits (0-9) in the tenths place (mm). We analyzed the first measurement from each anthropometrist. Since each child was measured by two different anthropometrists, we included 948 observations in the analysis. We used the Stata SE 13 (StataCorp, College Station, TX, USA) *digdis* package for analysis. We tested for any digit preference with Pearson's Chi Squared Test. We calculated the difference between the expected and observed percentages and tested for preference of each individual digit using a two-sided Binomial Test. To determine how close our observed proportions were to a uniform distribution, we summed the difference between expected and observed for positive differences only. The sum gives the percentage of digits that would need to be reclassified (moved from an overrepresented digit to an underrepresented digit) to achieve a uniform distribution, which is similar to the Myer's Blended Index (7) and the Digit Preference Score (14). We calculated the Digit Preference Score to allow comparisons with other research.

We used the 2006 WHO Growth Standards to calculate z-scores for weight-for-length/height (WHZ), length/height-for-age (HAZ), weight-for-age (WAZ), head circumference-for-age (HCZ), and arm circumference-for-age (ACZ). We assessed the plausibility of z-scores by determining the proportion of measurements flagged as falling outside of the plausible range as defined in WHO macros (WHZ <-5|>5; HAZ <-6|>6; WAZ <-6|>5; BMIZ <-5|>5; HCZ <-5|>5; ACZ <-5|>5) (15). We also examined flags used in Demographic and Health Surveys (DHS) for length and height plausibility (lying down 45-110 cm, standing up 65-120 cm). In the WHO Macro, implausible length or height measurements are not flagged, but WHZ scores that cannot be calculated are automatically set to missing. We calculated z-score standard deviations (SD) and analyzed SD disaggregated by age (under and over two years of age) to determine if the quality of measurements differed by age. We also analyzed z-scores for both the average of all measurements (repeated) and the first measurement of the first anthropometrist (single).

We analyzed both intra- and inter-observer error for height, HC and MUAC using repeated measurements. Intra-observer technical error of measurement (TEM) was calculated with the following formula:

$$TEMintra = \sqrt{\sum_{i=1}^{N}(M_{i1} - M_{i2})^2 / _{2 * N}}$$

, where N is the number of children and $M_{i1} M_{i2}$ are the closest repeated manual measures for one child by one observer. For inter-observer TEM we compared the average measurement of one observer to the average measurement of another observer for the same child, changing the numerator in the equation above to $(^{M_{ij1} + M_{ij2}}/_{2} - {}^{M_{ik1} + M_{ik2}}/_{2})^2$. We calculated Relative TEM, also known as %TEM, by dividing TEM by the mean of all the measurements that went into calculating TEM and multiplying by 100 (16). Total TEM combines intra and inter-observer reliability in the following formula:

$$TotalTEM = \sqrt{TEMintra^2 + TEMinter^2}$$

For correlation, we calculated the Intraclass Correlation Coefficient (ICC) using two-way mixed effects (17) and an absolute agreement definition in SPSS 20 (IBM Corp., Armonk, NY, USA).

3.3. Results from Quality Tests

3.3.1. Digit Preference

All measurements (stature, HC, and MUAC) showed evidence of terminal digit preference when we tested the child's first measurement with Pearson's Chi-Squared Test ($p<.01$, $n=948$). For all measurements, the terminal digit four was significantly overrepresented. In addition, eight was overrepresented for HC, and six was overrepresented for MUAC (Table 3-1). The sum of the difference between observed and expected percentages for all overrepresented digits indicated that 5.8% of stature measurements and 9.7% of both head and arm circumference measurements would need to be reclassified to achieve a uniform distribution (Table 3-1). The digit preference score was 5.5 for height, 8.7 for HC, and 8.3 for MUAC.

3.3.2. Plausibility & Reliability

Using means of repeated measures and ranges defined by WHO for biologically plausible z-scores one child (0.2%) was flagged, falling above the range for WHZ, BMIZ and ACZ. Also, the length of one child (0.2%) was lower than the plausible range. With single measures, the length of one additional child was lower than the plausible range, and no additional children had implausible z-scores.

For means of repeated measures, the standard deviation for all z-scores (HAZ, WHZ, WAZ, ACZ, and HCZ) was close to 1.0, ranging from 0.92 for WHZ to 1.07 for HAZ. Standard deviations using single measures were slightly higher, with the largest differences in ACZ (0.03) and WHZ (0.02) (Table 3-2). With increased variation we also found single measures had slightly higher prevalence estimates of children below or above some SD cutoffs for ACZ and WHZ (Figure 3-1), but differences were not statistically significant when tested with Chi-Square. For both repeated and single measures, z-score standard deviation was not consistently higher for children under 2 years of age compared to children 2 years and older. Levene's Test for Equality of Variances showed no statistical difference in z-score variance between the age groups for all indices (Table 3-2).

The correlation between an anthropometrist's first and second measurement of a child (intra-observer), and between two anthropometrists' average measurements of a child (inter-observer) was near perfect. The Intraclass Correlation Coefficient for all intra- and inter-observer measurements was exactly 1.00 (95% confidence interval of 1.00, 1.00) except for inter-observer arm circumference, with ICC of 0.99 (CI: 0.99, 0.99). Intra-observer TEM in centimeters was 0.22, 0.13, and 0.16 for stature, HC, and MUAC respectively. Intra-observer TEM values corresponded to relative TEMs of 0.26%, 0.29%, and 1.04% respectively (Figure 3-2). As an illustration of TEM interpretation, the stature intra-observer TEM of 0.22 cm means that 2/3^{rds} of repeated measurements were within ±0.22 cm and 95% of replicate measurements were within 2*TEM, or ±0.44

cm. Inter-observer TEM was higher than intra-observer for all measurements: stature TEM 0.34 cm, %TEM 0.42%; HC TEM 0.25 cm, %TEM 0.54%; and MUAC TEM 0.19 cm, %TEM 1.22% (Figure 3-2). Although MUAC inter and intra TEM was lower in absolute terms, MUAC relative TEM was approximately two to three times higher than stature and HC. Total TEM, combining inter- and intra-observer reliability, was 0.40 for stature, 0.28 for HC, and 0.25 for MUAC in centimeters.

3.4. Discussion

Our data shows that manual anthropometry in BINA was of excellent quality. Only 0.4% of the sample had biologically implausible z-scores or measurements, which is well below the 1% cutoff recommended by a WHO expert committee as an indicator of data quality problems (10). Our z-score SDs also indicated good quality because they were between 0.9 SD - 1.1 SD (14, 18), and there were no differences in SD between age groups; a higher SD among children under 2 is an indicator of poor quality and is attributed to difficulty in measuring the length of young, uncooperative children (7, 19). We found no terminal digit preference for zero or five, and the percent of measurements theoretically requiring reclassification for a normal distribution was low; we could obtain a normal distribution by changing 6% of terminal digits for our stature measurements compared to an average of 18% in 52 DHS from 2005-2014 (7). Biological plausibility, z-score SD, and digit preference are metrics commonly used to assess data quality of single anthropometric measures, and our study showed excellent quality according to all three metrics. In addition, the repeated measures in our study enabled analysis of inter- and intra-observer reliability. For stature the BINA intra- and inter-observer TEMs of 0.22 and 0.34 cm respectively are within the range of inter-observer TEM at MGRS study sites (0.15-0.41 cm).The TEMs for all of our measurements were on par with TEMs observed at MGRS study sites and were within the 95% precision margin of MGRS expert anthropometrists (4).The quality of manual anthropometry in BINA is similar to the quality of anthropometric data used to develop the 2006 WHO growth standards.

In our study, we achieved high quality manual anthropometry through following established protocols and adding additional quality control components. We followed advice from MGRS protocol authors to develop a study specific training manual, use proper, high quality equipment that is maintained, train and test staff with standardization, and provide adequate supervision during data collection (3). For measurement techniques, we drew from materials used in household surveys (20-23), which are largely based on a 1986 UN publication on measuring children (24). We also made sure that our techniques were aligned with the procedures used to develop the WHO Child Growth Standards by referring to MGRS protocol and the WHO anthropometry video used to train MGRS anthropometrists (3, 25). With the aim to strengthen their understanding of, and dedication to adhere to, study protocols, we included our field staff in development of the study manual. Field staff produced the first draft of the 3D imaging section of the manual and participated in revision and refinement of manual anthropometry sections. Field staff participation in study manual development led to recognition of the need to improve upon some of the borrowed materials. For example, the instructions for finding the mid-upper arm point did not adequately explain identification of the acromion process and illustrations incorrectly depicted measuring the side of the arm. For head circumference, we added additional instructions for the anthropometrists to reposition at both the front and back of the head while removing and replacing the tape multiple times to find the largest circumference. Electronic data capture allowed us to avoid data entry errors through the use of range checks and double data entry at the point of data acquisition; and facilitated adequate monitoring and supervision. Identifying errors during measurement allowed us to re-measure children, reducing the number of implausible measurements. Like the MGRS and following recommendations from a 1995 expert committee (3, 10), we routinely assessed field staff accuracy and provided regular, timely feedback on the quality of anthropometry. Assessment of accuracy during data collection is not standard in common surveys such as DHS and MICS. Repeated measurements, which are rare outside of research study settings allowed us to include reliability in quality monitoring; and electronic data capture provided anthropometrists with real-time feedback on their own reliability through the use of automatic triggers.

There have been suggestions to take repeated measures in surveys (11). Our results show little difference between single and repeated measures for z-score standard deviations and prevalence. However, differences may be greater with poor quality anthropometry, and repeated measures may influence quality. Triggering additional measurements provides an incentive to anthropometrists to improve reliability and avoid additional work. For survey institutions with a known history of poor quality anthropometry temporary adoption of duplicate measures by two anthropometrists, along with triggering of repeated measures based on inter-observer maximum allowable difference, may help to improve quality and provide accurate results. Inter-observer error was higher in our study, which is consistent with findings in a developing country setting (19); the MGRS used inter-observer triggering and adoption of inter-observer triggers in surveys may improve quality more than intra-observer triggers. In addition, repeated measures could be restricted to the most unreliable measures, such as length of children under two years of age and MUAC.

There is some consensus on what to assess for anthropometric data quality. A 1995 WHO expert committee recommended looking at accuracy, reliability, biological plausibility, z-score standard deviation, and digit preference (10); and these metrics are incorporated into manuals and assessments for DHS and MICS. SMART survey methodology includes all of these metrics, along with additional metrics such as skewness and kurtosis. The 2011 Emergency Nutrition Software (ENA) for SMART has automated reports for standardization tests and data quality, with the latter referred to as a plausibility check (14). While quality assessments for DHS, MICS and SMART are similar, the methods are not exactly the same. For example, z-score ranges for biological plausibility can differ and there are multiple tests for digit preference. Our study provides a simplified method for assessing digit preference that does not require specialized software. Standardization of what to measure and how to measure anthropometric data quality is needed, but there may be larger institutional differences in how to interpret and act on data quality reports. One of the challenges in this study was determining the criteria to judge whether or not we achieved gold standard manual anthropometry. DHS and MICS do not have indicator thresholds for acceptable anthropometric data quality, and it is very rare for either to suppress data because of poor quality. The ENA SMART plausibility check determines if data quality is acceptable with a composite score based on thresholds for multiple indicators (14). Another composite scoring system was developed by UNICEF-supported research (9), but the system is not regularly used. For many surveys data quality is assessed, but there is no standard criteria to judge whether or not results should be released.

We can consider z-score standard deviation to illustrate the importance of reaching consensus on interpretation and action. WHO and the US CDC promote the use of normative ranges of SD to determine if survey quality is acceptable (10, 26), but the ranges are based on surveys that have evidence of poor data quality (7, 18). The most recent DHS data quality assessment showed that 30 of 52 countries had HAZ SD greater than 1.5, but only one country suppressed data because of poor quality (7). According to SMART data quality is not acceptable if HAZ SD is above 1.2 (14), and a recent modeling study showed that SD of 1.5 can result in substantial overestimation of stunting prevalence (18). Meanwhile, the published normative range for HAZ SD that some organizations use to deem data quality acceptable is 1.35-1.95 (10, 26). Digit preference provides another example of the need to focus on interpretation and action. Digit preference has been monitored for years as an indicator of data quality, but even very high measurement digit preference (80% theoretically needing reclassification for a normal distribution) has no meaningful impact on z-score means or prevalences (18). Weight and height heaping may be useful as a proxy for measuring whether or not anthropometrists follow protocol, but the indicator has little bearing on if and how data should be used. The ENA SMART plausibility check provides a good starting point for consensus. Efforts for standardization should consider that quality indicator thresholds should be based not only on what is feasible, as was done in the MGRS for reliability and accuracy (3), but also on how quality affects the usefulness of individual growth monitoring, prevalence estimates, and population-level trend analysis.

3.5. Conclusions

Manual anthropometry in BINA should be considered gold standard, and the high quality can likely be attributed to highly motivated and competent field staff, reduction of non-measurement error through the use of technology, and reduction of measurement error through adequate monitoring and supervision. We made slight improvements to training materials and added to the MGRS protocol for quality control by utilizing electronic data capture and including field staff in the development of the study manual. BINA methodology can be followed to improve anthropometric data quality, but will need to be adapted for use in some contexts. For example, BINA field staff were highly educated and reported child age is accurate in the US; in a different context different strategies may be needed to allow field staff to contribute to manual development, and double data entry may not be sufficient to ensure quality age data. Furthermore, for systematic improvement of anthropometric data in clinics and surveys the availability of adequate methodology may not be enough. The lack of consensus on how to interpret and act on analysis of anthropometric data quality leads to results being published without reproach, which limits the institutional need and motivation to improve quality. Ultimately, while high quality anthropometry is possible with available manual equipment, improved technology may be the most efficient driver of widespread quality improvement; and BINA can provide a good evaluation of 3D imaging for child anthropometry because manual measurements were of excellent quality.

3.6. Figures and Tables

Table 3-1 Terminal digit preference expressed as percentage of anthropometrists' first measurement for height, head circumference and arm circumference ending in .0 to .9 compared to the expected 10% among children 0-4.9 years (n=948 observations from 474 children), BINA 2017

Terminal Digit (tenths place)	Height			Head Circumference			Arm Circumference		
	% Observed	% Observed - % Expected	P-Value	% Observed	% Observed - % Expected	P-Value	% Observed	% Observed - % Expected	P-Value
0	8.5	-1.5	0.14	5.0	-5.0	0.00	5.9	-4.1	0.00
1	9.2	-0.8	0.45	9.2	-0.8	0.45	10.7	0.7	0.48
2	11.2	1.2	0.23	11.2	1.2	0.23	8.4	-1.6	0.12
3	10.2	0.2	0.79	10.1	0.1	0.87	9.8	-0.2	0.91
4	14.3	4.3	0.00	14.3	4.3	0.00	14.6	4.6	0.00
5	9.7	-0.3	0.83	6.8	-3.2	0.00	6.4	-3.6	0.00
6	9.8	-0.2	0.91	9.8	-0.2	0.91	12.3	2.3	0.02
7	9.8	-0.2	0.91	9.6	-0.4	0.75	11.6	1.6	0.10
8	9.1	-0.9	0.36	13.3	3.3	0.00	9.7	-0.3	0.83
9	8.1	-1.9	0.06	10.8	0.8	0.42	10.5	0.5	0.55

Table 3-2 Mean z-score, standard deviation, and test of equal variance between age groups for height-for-age (HAZ), weight-for-height (WHZ), weight-for-age (WAZ), arm circumference-for-age (ACZ), and head circumference-for-age (HCZ) from manual measurements (n=474), BINA 2017

		Total (0-4.9 years)			Less than 2 years (U2)			2-4.9 years (O2)			Difference in Variance (Levene's Test)		
		n	Mean z-score	Standard Deviation (SD)	n	Mean z-score	Standard Deviation	n	Mean z-score	Standard Deviation	U2 SD-O2 SD	F	P-Value
Repeated measure mean	Height-for-age	474	-0.29	1.07	223	-0.42	1.10	251	-0.18	1.03	0.08	0.12	0.73
	Weight-for-height	472	0.34	0.92	222	0.32	0.88	250	0.35	0.96	-0.08	0.87	0.35
	Weight-for-age	474	0.06	1.04	223	-0.05	1.02	251	0.17	1.05	-0.04	0.26	0.61
	Arm circumference-for-age	385	0.78	0.94	135	0.84	0.92	250	0.75	0.95	-0.03	<0.01	0.93
	Head circumference-for-age	474	0.24	1.02	223	0.11	1.01	251	0.35	1.01	0.00	0.32	0.57
Single measure	Height-for-age	474	-0.30	1.08	223	-0.43	1.11	251	-0.17	1.02	0.09	<0.01	0.96
	Weight-for-height	471	0.34	0.95	221	0.34	0.92	250	0.34	0.97	-0.05	0.38	0.54
	Weight-for-age	474	0.06	1.04	223	-0.05	1.01	251	0.17	1.05	-0.04	0.31	0.58
	Arm circumference-for-age	385	0.78	0.97	135	0.85	0.98	250	0.75	0.96	0.02	0.25	0.62
	Head circumference-for-age	474	0.24	1.04	223	0.11	1.04	251	0.36	1.03	0.01	0.52	0.47

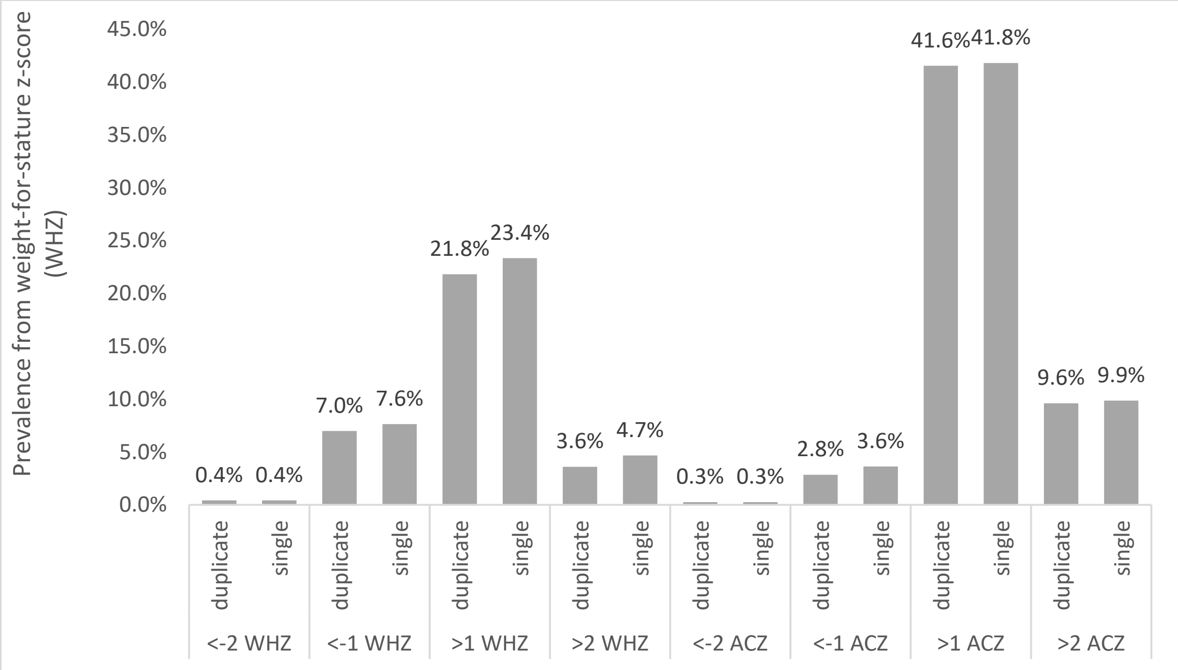


Figure 3-1 Prevalence by standard deviation cutoffs for weight-for-height and arm circumference-for-age z-scores for single and means of repeated measures among children 0-4.9 years of age, BINA 2017

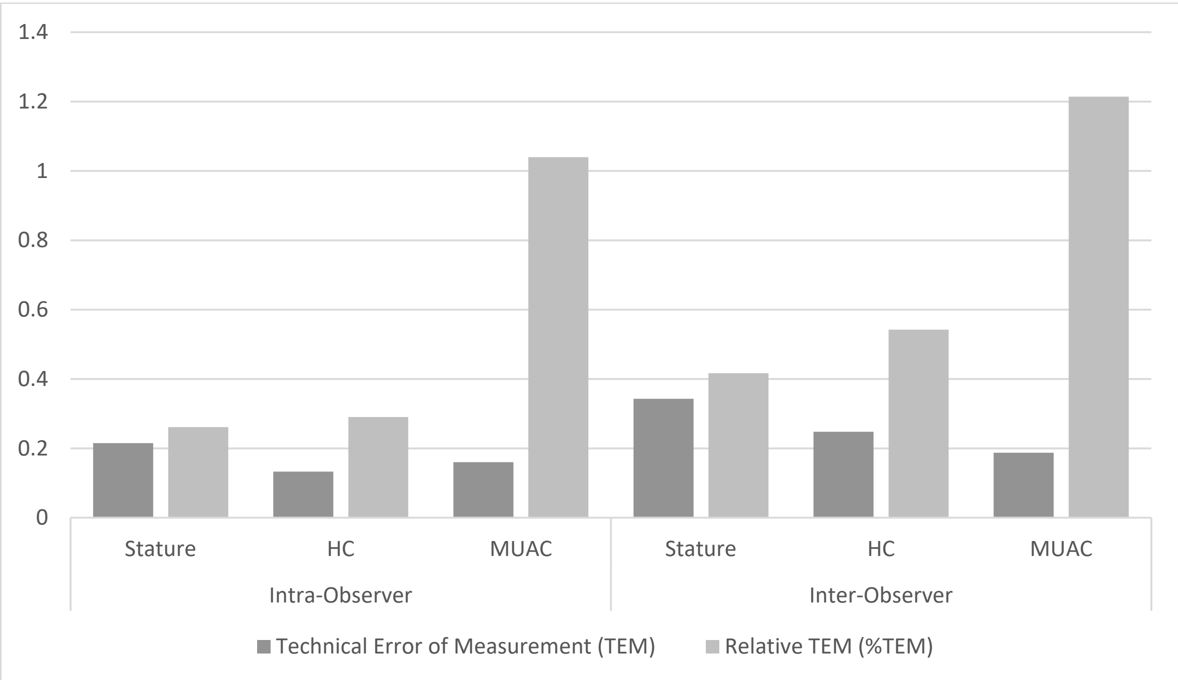


Figure 3-2 Measurement reliability in closest two manual measures from single observer (intra-observer) and measurement reliability in average of closest two manual measures between two observers (inter-observer) for stature, head circumference (HC) and arm circumference (MUAC) among children 0-4.9 years of age (intra-observer n=948, inter-observer n=474), BINA 2017

References (Chapter 3)

1. De Onis M, World Health Organization. Dept. of Nutrition for Health and Development. WHO child growth standards : length/height-for-age, weight-for-age, weight-for-length, weight -for-height and body mass index-for-age : methods and development. Geneva: World Health Organization; 2006.

2. Wit JM, Himes JH, van Buuren S, Denno DM, Suchdev PS. Practical Application of Linear Growth Measurements in Clinical Research in Low- and Middle-Income Countries. Horm Res Paediatr. 2017.

3. de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr Bull. 2004;25(1 Suppl):S27-36.

4. WHO Multicentre Growth Reference Study Group. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl. 2006;450:38-46.

5. Yin H, Dai Y, Li H, Xie X, Ren H. The test–re-test reliability of routine infant anthropometry at primary care hospitals in Chongqing, PR China. Annals of Human Biology. 2013;40(4):309-17.

6. Gerner B, McCallum Z, Sheehan J, Harris C, Wake M. Are general practitioners equipped to detect child overweight/obesity? Survey and audit. J Paediatr Child Health. 2006;42(4):206-11.

7. Assaf S, Kothari, M., Pullum, T. An Assessment of the Quality of DHS Anthropometric Data, 2005-2014. DHS Methodological Reports No 16. Rockville, MD, USA: ICF International; 2015.

8. Pullum TW. An Assessment of the Quality of Data on Health and Nutrition in the DHS Surveys, 1993-2003. Methodological Reports No 6. Calverton, MD, USA: Macro International Inc.; 2008.

9. Corsi DJ, Perkins JM, Subramanian SV. Child anthropometry data quality from Demographic and Health Surveys, Multiple Indicator Cluster Surveys, and National Nutrition Surveys in the West Central Africa region: are we comparing apples and oranges? Glob Health Action. 2017;10(1):1328185.

10. WHO Expert Committee on Physical Status. The Use and Interpretation of Anthropometry. WHO Technical Report Series 854. Geneva: World Health Organization; 1995.

11. USAID. Anthropometric Data in Population-Based Surveys, Meeting Report, July 14-15, 2015. Washington, D.C., USA: FHI 360/FANTA; 2016.

12. SMART. ENA Software for SMART: ACF; 2015 [Available from: http://smartmethodology.org/survey-planning-tools/smart-emergency-nutrition-assessment/.

13. US CDC International Micronutrient Malnutrition Prevention and Control Program. Micronutrient Survey Toolkit Toronto: Micronutrient Initiative; 2016 [Available from: http://surveytoolkit.micronutrient.org/.

14. SMART Action Against Hunger - Canada and Technical Advisory Group. The SMART Plausibility Check for Anthropometry. Toronto; 2015.

15. WHO. WHO Anthro and macros Geneva: WHO; 2015 [Available from: http://www.who.int/childgrowth/software/en/.

16. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. The British journal of nutrition. 1999;82(3):165-77.

17. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. Stat Med. 1994;13(23-24):2465-76.

18. Grellety E, Golden MH. The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition. PloS one. 2016;11(12):e0168585.

19. Ayele B, Aemere A, Gebre T, Tadesse Z, Stoller NE, See CW, et al. Reliability of measurements performed by community-drawn anthropometrists from rural Ethiopia. PloS one. 2012;7(1):e30345.

20. Cogill B. Anthropometric Indicators Measurement Guide. Washington, D.C.: Food and Nutrition Technical Assistance (FANTA) Project, FHI 360; 2003.

21. US CDC. National Health and Nutrition Examination Survey: Anthropometry Procedures Manual. Atlanta, GA, USA: US CDC; 2007.

22. UNICEF. Multiple Indicator Cluster Surveys: Manual for Anthropometry. New York, NY, USA: UNICEF; 2014.

23. ICF International. MEASURE DHS Biomarker Field Manual. Calverton, MD, USA: ICF International; 2012.

24. United Nations. Department of Technical Co-operation for Development and Statistical Office., National Household Survey Capability Programme. How to weigh and measure children : assessing the nutritional status of young children in

household surveys. Preliminary ed. New York: United Nations; 1986. 94 p. + Annex in pocket. p.

25. World Health Organization. Training Course on Child Growth Assessment: WHO Child Growth Standards. Geneva: WHO; 2008.

26. Mei Z, Grummer-Strawn LM. Standard deviation of anthropometric Z-scores as a data quality assessment tool using the 2006 WHO growth standards: a cross country analysis. Bull World Health Organ. 2007;85(6):441-8.

Chapter 4 . Accuracy and Reliability of a Low-Cost, Handheld 3D Imaging System for Child Anthropometry

Joel Conkle, MPH[a*], Parminder S. Suchdev, MD MPH [a,b,c,d], Eugene Alexander, PhD[e] , Rafael Flores-Ayala, DrPH[a,c], Usha Ramakrishnan, PhD [a,b], and Reynaldo Martorell, PhD [a b]

[a] Doctoral Program in Nutrition and Health Sciences. Laney Graduate School. Emory University. Atlanta, GA, USA; joel.conkle@emory.edu
[b] Hubert Department of Global Health. Rollins School of Public Health. Emory University. Atlanta, GA, USA
[c] Division of Nutrition, Physical Activity and Obesity. National Center for Chronic Disease Prevention and Health Promotion. U.S. Centers for Disease Control and Prevention. Atlanta, GA, USA
[d] Department of Pediatrics, School of Medicine, Emory University. Atlanta, GA, USA
[e] Body Surface Translations, Inc.. Atlanta, GA, USA
* Correspondence: joelconkle@gmail.com; Tel.: +1-404-725-7138

Abstract: Anthropometry is used clinically to identify and manage malnutrition, and in national surveys to guide nutrition programs. The usefulness of anthropometry, however, is undermined by poor measurement quality, which has led to calls for improved quality and new measurement approaches. We evaluated the ability of a three-dimensional (3D) imaging system to correctly measure child stature (length or height), head circumference (HC) and mid-upper arm circumference (MUAC).

We manually measured 474 apparently healthy children 0-5 years of age in Atlanta, USA following techniques used to develop the 2006 WHO Child Growth Standards. We took 3D scans of the same children using an iPad Air® with Structure Sensor® and derived measurements from scans using AutoAnthro, custom software developed by Body Surface Translations. We evaluated the reliability and accuracy of 3D scan-derived measurements against manual measurements.

Measurement reliability of repeated scans was within 1 mm of manual measurement reliability for stature, HC and MUAC. We found systematic bias when analyzing accuracy — on average 3D imaging overestimated stature and HC by 6 mm and 3 mm respectively, and underestimated MUAC by 2 mm. After adjusting measurements to remove systematic bias, 3D imaging yielded mean z-scores, z-score standard deviations (SD), and prevalence below or above z-score SD cutoffs that were similar to manual measurements ($p>0.1$). Based on a cutoff of one SD, specificity of adjusted, scan-derived measurements was above 0.95 for all measures, and sensitivity was above 0.90 for stature and MUAC. HC sensitivity was 0.87.

The 3D imaging system used in this study is low-cost, portable and can handle movement, making it ideal for use in the field. However, additional research, particularly on accuracy, and further development of the scanning and processing software is needed before making policy and practice recommendations on the use of 3D imaging for child anthropometry. The Bill and Melinda Gates Foundation funded the study.

4.1. Introduction

Body measurement, or anthropometry, can be compared to a reference population to define nutritional status and to monitor child growth. Length or height, weight, and head circumference (HC) are common measures for infants and children under 5 years of age. Anthropometry is used clinically to diagnose malnutrition (1-4), to identify underlying conditions (3), to assess risk for future disease (5, 6), and for clinical research (7). At the population level, public health practitioners include anthropometry in research and surveys to identify causes and effects of abnormal nutritional status, to monitor trends through surveillance, and to target and evaluate interventions related to nutrition (6). Anthropometry is also used to evaluate agricultural initiatives, and the global development community uses population-level anthropometry as an indicator of national economic development. Height-for-age is accepted as a more comprehensive indicator of poverty than income, (8) and there is recognition that nutrition is essential for human capital development (9). There is a target to improve stunting in the Sustainable Development Goals, (10) and anthropometric indicators are used for allocation of Official Development Assistance (11).

Given that child growth has broad effects on health, nutrition, and development, it is important that anthropometric measurements are of high quality. Studies in primary care facilities of developed countries found that measurement error led to inaccurate and unreliable circumference measurement for adults (12, 13) and unreliable length and circumference measurements for children (14, 15). There is also evidence that a lack of standardization and maintenance of anthropometric equipment in health facilities leads to misclassification of child weight status (16). Three separate evaluations covering hundreds of large-scale, established surveys in developing countries found that on average more than 3% of weight or height measurements were biologically implausible (17-19). According to a WHO Expert Committee, when more than one percent of measurements are considered biologically implausible, a survey is likely to be of poor quality (20).

The usefulness of anthropometry is undermined by poor measurement quality, which has led to calls for the use of technology to improve quality of child anthropometry (17, 21). This study evaluated the ability of a portable, three-dimensional (3D) imaging system to accurately and reliably measure child stature (length or height), head circumference, and mid-upper arm circumference (MUAC).

4.2. Materials and Methods

4.2.1. Study Design and Participants

We designed the Body Imaging for Nutritional Assessment Study (BINA) to evaluate the accuracy and reliability of a 3D imaging system in comparison to manual measurements for child anthropometry. The study was approved by the Emory Institutional Review Board, and included two phases. In the first phase we calibrated software to process 3D scans into measurements by scanning and measuring 36 children. In the second phase, the topic of this paper, we tested 3D imaging on a new sample of children. Children 0-5 years of age who were apparently healthy and whose primary caregiver gave informed, written consent were eligible for the study.

Caretakers received a $15 gift card for each child participating in the study. We recruited and measured children at 20 facilities in and around metro Atlanta, GA, USA; including at daycare, higher education, religious, and medical facilities. We selected recruitment sites to reflect a generally representative population of Atlanta children and included a maternity ward to oversample newborns. Daycare centers received gift cards for participating as a study site. We formed a convenience sample by recruiting children on-site, via email, and through facility administrative staff; recruitment was ongoing throughout data collection, which lasted from September 2016 to February 2017. The intended sample size for the study was 500.

4.2.2. Test Methods

Five trained anthropometrists with post-secondary education performed all manual measurements and 3D scans. Anthropometrists received training over a three week period in August 2016 from expert anthropometrists at Emory University and passed a standardization test for manual anthropometry. Manual measurements followed the protocol used to develop the 2006 World Health Organization (WHO) Child Growth Standards (CGS) (22); detailed methods for manual anthropometry in BINA are published elsewhere (23). Staff from Body Surface Translations (BST) trained anthropometrists to take 3D scans in one day, and anthropometrists informally used 3D scanners throughout the three week training period to familiarize themselves with the technology. During the standardization test anthropometrists scanned children following study protocol, and after visual assessment we determined scans were of sufficient quality to proceed with the study.

Each anthropometrist carried a 3D scanning device, an iPad Air® (Apple, Cupertino, CA) with attached Structure Sensor (Occipital, San Francisco, CA, USA), and custom software from BST, AutoAnthro, for scanning and data entry of demographic information and manual measurements. We collected scans and then manual measurements consecutively at the same time of the day, usually in the morning. Each individual 3D imaging session comprised six scans, with three scans of the front of the child and three of the back. For each scan of children two years of age and older an assistant positioned the child's arms in one of three poses (Figure S1); for children less than two years of age we positioned arms away from the torso, but not necessarily in the three identified poses. An individual scan captured 30 frames over one second. Multiple scans for one session provided redundancy in the case that movement blurred or hid a part of the child's body. Consistent with manual anthropometry procedures, we scanned children two years of age and over standing up, and instructed younger children to lie down. Children were measured undressed to their diaper or in skin-tight shorts/leotards provided by the study. Each child was scanned and measured twice by two different people, resulting in four sessions of scans and four sessions of manual measurements per child. Multiple measurements allowed analysis of both inter- and intra-measurer reliability. Both members of a field team separately entered the child's name, sex, race, ethnicity, and birth date from consent forms previously filled out by the child's caretaker. AutoAnthro automatically uploaded data, including scans, to an online database. There were no reported adverse events from scans or manual measurements.

Anthropometrists received regular supervisor feedback on accuracy and reliability of manual measurements. Software included range checks to avoid data entry error for manual anthropometry, and automatic triggering of a third manual measurement based on Maximum Allowable Difference (22) to improve reliability. For demographic information we cleaned the data by checking for inconsistencies in double data entry. Following data collection we processed scans to produce scan-derived measurements with no consideration of manual measurements. AutoAnthro had automated processing, which included automatically removing the image background and fitting the 3D point cloud obtained from each scan (Figure S2) to an animator's model, producing an intermediate result of six fitted models. For final measurements, AutoAnthro combined the six fitted models into a single, final model with best fit; and derived measurements based on points pre-selected once on the base model (Figures S3 and S4). We used two separate animator's models: one for children under one month and another for children 1-59.9 months. To improve fitting efficiency, we used median measurement-for-age from WHO CGS (24) to automatically scale the animator's model prior to fitting.

4.2.3. Analysis

In this study, one anthropometrist could be triggered to take a third measurement for manual measurements, but not for scans. To determine a best-estimate from manual measurements, we excluded the outlying measurement in the case of a triggered, third measurement; and took the mean from the four remaining measurements (two from each anthropometrist). In this paper we refer to the average of four manual measurements as the "best-estimate," and consider the best estimate the reference standard. For analyzing reliability we limited our analysis to the first two manual measurements, ignoring any triggered third measurement; which provided a like-for-like comparison with scan-derived measurements. In the text we refer to the mean of two manual measurements as "repeated-manual," and to measurements derived from one manual measurement as "single-manual." We analyzed scan-derived measurements using different sets of scans. We used scan-derived measurements based on a single session (6 scans from a single observer), repeated sessions (average from two sessions of a single observer), and all sessions (mean of four sessions from two observers); referred to as "single-scan", "repeated-scan", and "all-scan" in the text. For single-manual and single-scan, we used the first observation from the first measurer for each child. For repeated-manual and repeated-scan we used both observations from the first observer.

We analyzed accuracy, reliability, z-scores and classification bias using SPSS 20 (IBM Corp., Armonk, NY, USA) and StataSE 13 (StataCorp, College Station, TX, USA). For accuracy we considered both systematic and random bias, evaluating Average Bias and Bland-Altman (BA) Plots (25). We used Technical Error of Measurement and the Coefficient of Reliability as described by Ulijaszek (26) to measure reliability, and for classification bias we analyzed individual level sensitivity and specificity. Detailed methods for analysis are included in supplementary online text.

4.3. Results

4.3.1. Participation and Sample Characteristics

Figure S5 shows the flow of participants in the study. We received informed consent for 555 children, of which 26 children were either not present or had aged out by the day of data collection. Of the remaining 529, we excluded 55 due to: refusal to be measured (n=18), incomplete measurements (n=8), health status (n=5), loss of data due to technical errors during upload (since corrected) (n=10), and use of child in calibration of the 3D imaging system (n=14); resulting in a final sample size of 474. Table 4-1 presents sample characteristics. There was racial and ethnic variation in the sample, and approximately equal numbers of boys and girls. A total of 47% of the final sample was under two years of age; newborns were overrepresented, and nearly all of the newborns were less than four days old. There was a low prevalence of wasting, stunting, underweight and overweight.

4.3.2. Accuracy

Table S1 shows the accuracy of scan-derived measurements when compared to best-estimate manual measurements. The average bias of scan-derived measurements in cm was +0.6 for stature, +0.3 for HC, and -0.2 for MUAC; and differences were consistent whether measurements were derived from single-scan, repeated-scan, or all-scan. However, the number of scan sessions did have an effect on the spread of differences. Averaging repeated measurements reduced variance and narrowed the spread of differences as expected. For stature 97% of all-scan measurements were higher than manual measurements, or positive, and the 95% limit of agreement (LoA) showed that 95% of individual differences were within -0.1 to 1.2 cm; single-scan measurements were 78% positive with a LoA of -0.7 to 1.9 cm.

Bland-Altman Plots and related statistics are presented in Figure 1. Compared to children 1-59.9 months of age 3D imaging was less accurate for newborns for all measures (stature, HC, and MUAC). When considering all children under five years of age, there was evidence that accuracy of scan-derived measurements changed by the size of the child for all measures. After disaggregating by age group (corresponding to the two animator's models) Pitman's Test was no longer statistically significant for stature and HC, indicating that accuracy was different between the two age groups, but that there was not differential accuracy by size within the two age groups for both measures. For MUAC, the Pitman's Test remained statistically significant after disaggregating by age group, suggesting differential accuracy by size within both age groups. Subsequent linear regression of the difference between scan-derived and manual measurements on an independent measure of size confirmed differential accuracy by size for MUAC — 3D imaging was less accurate for children with smaller MUAC. After separating children 1-59.9 months of age into quintiles based on MUAC, average bias of scan-derived measurements in cm was -0.31 (MUAC 9.6-15.1 cm), -0.18 (MUAC 15.1-16.0 cm), -0.15 (MUAC 16.0-16.7 cm), -0.02 (MUAC 16.7-17.6 cm), and -0.05 (MUAC 17.6-25.3 cm).

Among children 1-59.9 months of age there were no statistically significant or meaningful differences in accuracy by race or hairstyle (Table S2). The largest difference was a 0.04 cm difference in average bias for head circumference between Black and White children.

4.3.3. Reliability

Table S3 displays inter- and intra-observer TEM, %TEM and ICC for scan-derived and manual measurements. TEM represents one standard deviation and a 95% precision margin can be calculated by multiplying TEM by two. The intra-observer TEM for stature among children of all ages was 0.62 cm for scan-derived measurements, indicating that for a single observer the second scan-derived stature was within ±0.62 cm of the first scan-derived stature for ~two out of three children, and that for 95% of children the difference was within ±1.2 cm. Manual measurement intra-observer TEM for stature among children of all ages was within ±0.72 cm for 95% of children. Intra-observer TEM from scan-derived measurements was higher than that from manual measurements for all measures and across all age groups (Figure 2a). For stature and head circumference, the highest manual measurement intra-observer TEM was among children 12-23.9 months of age, while children over three years of age had the lowest TEM. For intra-observer TEM based on scan-derived measurements, there were no meaningful differences by age group (Figure 2a).

Inter-observer reliability in Table S3 compares the average from repeated measurements by one observer to the average of another observer. Inter-observer TEM from scan-derived measurements was similar to inter-observer TEM from manual measurements across all measures and most age groups (Figure 2b). For all children under 5 years of age inter-observer TEM from repeated scans was within 0.1 cm of TEM from repeated manual measurements for all measures. We also looked at inter-observer TEM based on single measurements. Single-scan inter-observer TEM was higher than single-manual inter-observer TEM (Figure 3). When using single measurements inter-observer TEM was higher than intra-observer TEM for manual measurement, but not for scans (Figure 3).

Total TEM combines the intra- and inter-observer TEM from Table 4 into a single metric. For manual measurements Total TEM was 0.51 cm, 0.33 cm, and 0.31 cm for stature, HC and MUAC respectively; compared to 0.77 cm, 0.51 cm, and 0.43 cm for scan-derived measurements. Reliability coefficients based on Total TEM were 1.00, 1.00, and 0.99 for stature, HC and MUAC from manual measurements; and 1.00, 0.99, and 0.98 for scan-derived measurements.

4.3.4. Z-Scores and Classification

For scan-derived measurements we calculated z-scores based on both unadjusted and adjusted measurements, with adjustments made to remove systematic inaccuracy by subtracting or adding the average bias in Figure 1 from/to each observation. All height-for-age z-scores (HAZ) and head circumference-for-age z-scores (HCZ) were biologically plausible. One child was flagged for an implausibly high arm-circumference-for-age z-score (ACZ), and this same child was flagged by both manual and scan-derived measurements (with or without adjustment). Mean z-score values and percentages below or above SD cutoffs presented in Table 2 show that, as expected, the accuracy adjustments helped to reduce the difference between scan-derived and best-estimate results. Unadjusted, single-scan HC mean z-score was 0.2 higher than the best-estimate mean z-score, and the percentages above 1 SD and 2 SD were overestimated by five percentage points; after adjustment, single-scan HC and best-estimate mean z-scores were essentially the same and percentages differed by less than two percentage points (Table 2). Percentages based on adjusted, scan-derived measurements from single or repeated sessions were within two percentage points of percentages based on best-estimate manual measurements for all measures (Table 2). Z-score SDs for single-scan were higher than z-score SDs from single-manual (Table 2), which is the effect of less reliability seen in Figure 3. Using repeated-scan brought the z-score SD closer to the level of manual measurements.

For individual level agreement we analyzed sensitivity and specificity of HAZ <-1 SD, HCZ>1 SD, and ACZ>1 SD among children 1-59.9 months with the best-estimate considered "true" nutritional status. Sensitivity for stature measured the probability that a child's HAZ based on the adjusted scan-derived measurement was <-1 SD given that HAZ based on the best-estimate was <-1 SD. Stature sensitivity was 0.95, 0.92, and 0.93 for single-manual, single-scan, and repeated-scan respectively; indicating that out of 100 children identified as HAZ <-1 SD by the best-estimate, 95, 92 and 93 children would also be identified by single-manual, single-scan, and repeated-scan respectively (Table 3). Stature specificity was also high; out of 100 children not identified as HAZ <-1 SD, 98, 96, and 97 of those children would also not be identified as having HAZ <-1 SD by single-manual, single-scan, and repeated-scan respectively. For stature and arm circumference, sensitivity and specificity of single- and repeated- scan was excellent and performed nearly as well as single-manual. For scan-derived head circumference specificity was excellent, but sensitivity was 0.84 and 0.87 for measurements based on single-scan and repeated-scan respectively (Table 3).

4.4. Discussion

In a previous publication we concluded that BINA collected gold-standard, manual anthropometry based on analysis of biological plausibility, reliability, and z-score standard deviations (23). In this paper we compared measurements derived from 3D imaging to our gold-standard, manual measurements. For biological plausibility 3D imaging and manual measurement were exactly the same, with both methods producing plausible measurements >99% of the time; this finding indicates acceptable quality based on WHO expert committee criteria for biological plausibility (20). We also found that repeated-scan 3D imaging produced measurement reliability that was within 1 mm of manual measurement reliability for stature, HC and MUAC; a level of reliability that puts 3D imaging on par with manual measurements collected in the Multicenter Growth Reference Study (MGRS) used to develop the 2006 WHO CGS (23, 27). Considering only biological plausibility and reliability, 3D imaging performed as well as gold-standard manual measurements for child anthropometry. However, 3D imaging systematically underestimated or overestimated child size when compared to our best-estimate of size from manual measurement. We were able to obtain accurate results from 3D imaging that yielded mean z-scores and z-score standard deviations similar to manual measurements, but only after making adjustments to remove systematic inaccuracy. After adjustment the percentages of the population below or above z-score cutoffs for HAZ, HCZ, and ACZ according to 3D imaging were similar to manual measurements, and there was a high probability that a child identified as being below or above z-score cutoffs by manual measurement would be identified the same way by 3D imaging. Anthropometry is used to classify the nutritional status of individuals and populations. There is no single standard for judging the adequacy of child anthropometry reliability and accuracy (23), but ultimately if a method correctly classifies nutritional status and can detect changes in nutritional status over time, reliability and accuracy of that method should be considered acceptable. In our study 3D imaging was reliable enough to be considered a good method for collecting child anthropometry, but systematic inaccuracy caused misclassification. Additional discussion on reliability, and an examination of our findings in relation to monitoring and classification of nutritional status is included in supplementary online text.

Before reaching any conclusion on the readiness of 3D imaging for child anthropometry, we would need to determine if the systematic inaccuracy found in this study is population specific. If the same under- and overestimation was found in a different sample with different anthropometrists, we could then identify and fix the cause of the bias in the model fit or simply build adjustments into the software. Human error in manual measurement of MUAC and scan processing software that did not account for the soft tissue compression and body positioning specified in manual measurement protocol could have caused the systematic inaccuracy found in this study. Research similar to BINA should be carried out, ideally in developed and low and middle income countries, to help answer questions on systematic inaccuracy and also to address some of the other limitations of our study. The scanners used in our 3D imaging system do not function well in direct sunlight, which was not a constraint in BINA because we took scans inside of facilities. The 3D imaging system may perform differently in the context of a household survey or during community-based screening. Additional limitations to our study stem from sampling design, the nutritional status of our sample, and automated processing. Our sample was well-nourished; we did not have a sufficient number of children with abnormal nutritional status to analyze prevalence or sensitivity/specificity for clinically significant indicators, such as obesity, wasting and severe stunting. In addition, our sample was not random, age was not normally distributed, and findings cannot be generalized to any specific age group. The processing of 3D scans was not fully automated as planned; anthropometrists took more scans than needed and manually selected the best quality scans, and the orientation (front/back) of each scan was manually coded. Further software development is needed to achieve full automation. Additional discussion, including detailed hypotheses for the causes of bias, is included in supplementary online text.

3D imaging is not new for anthropometry (28-32), but the system used in our study was inexpensive and brought unique functionality. The scanning device we used is off-the-shelf, commercial hardware; and is a fraction of the cost of other systems (~$878 for hardware making up single scanner). The scanning device is small, lightweight, and the software developed by BST only requires a series of snapshots, which allows some subject movement. The 3D imaging system used in our study, AutoAnthro, is an ideal replacement for bulky height boards used in surveys, and to our knowledge it is the first portable 3D system specifically designed for whole body scanning of infants and young children. In conclusion, our findings indicate that AutoAnthro can produce reliable child anthropometry, but further research and development is needed before 3D imaging can be recommended as a solution to improving the quality of anthropometric data.

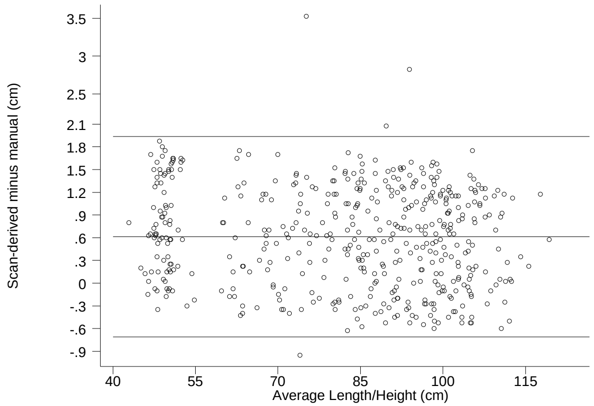
4.4. Figures and Tables

Table 4-1 Sample characteristics, BINA 2017

Age in months, mean (range)	25.7	(0-59)
Age Groups		
Newborn (<1 month)	82	(17%)
1-11.9 months	66	(14%)
1-1.9 years	75	(16%)
2-2.9 years	85	(18%)
3-4.9 years	166	(35%)
Sex		
Female	228	(48%)
Race		
Black	201	(42%)
White	134	(28%)
Asian	40	(8%)
Multiple, Other or Not Reported	99	(21%)
Ethnicity		
Non-Hispanic	385	(81%)
Hispanic	77	(16%)
Not Reported	12	(3%)
Anthropometric Indices		
Weight-for-Age Z-score (WAZ), mean ± SD	0.06	1.04
Height-for-Age Z-score (HAZ), mean ± SD	-0.29	1.07
Weight-for-Height Z-score (WHZ), mean ± SD	0.34	0.92
Head Circumference Z-Score (HCZ), mean ± SD	0.24	1.02
Arm Circumference Z-Score (ACZ), mean ± SD	0.78	0.94
Nutritional Status		
Underweight (<-2 SD WAZ)	11	(2.3%)
Stunted (<-2 SD HAZ)	21	(4.4%)
Wasted (<-2 SD WHZ)	2	(0.4%)
Overweight (>2 SD WHZ)	22	(4.7%)

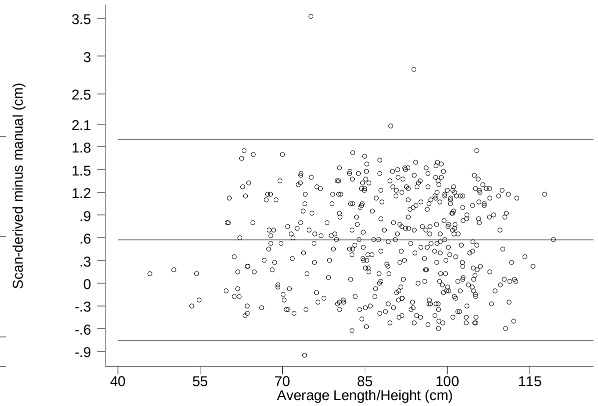
0-4.9 years of age	≥1 month of age	<1 month of age

A.



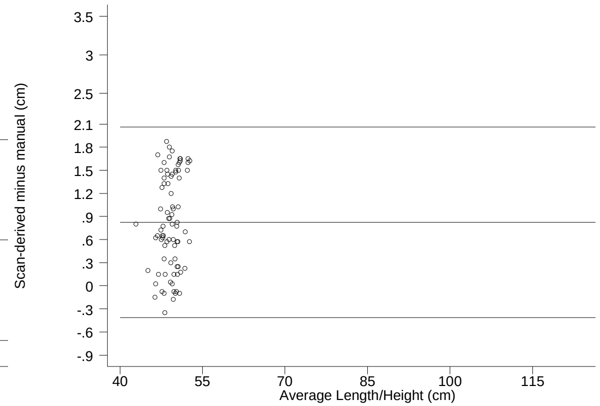
Pitman's Test: r = -0.112, n = 474, p = 0.015
Limits of agreement: -0.710 to 1.939
Mean difference: 0.615 (CI 0.555 to 0.674)

B.



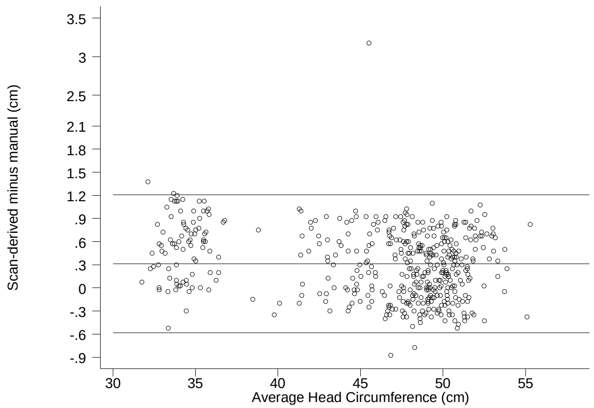
Pitman's Test: r = -0.005, n = 392, p = 0.919
Limits of agreement: -0.756 to 1.897
Mean difference: 0.571 (CI 0.505 to 0.636)

C.



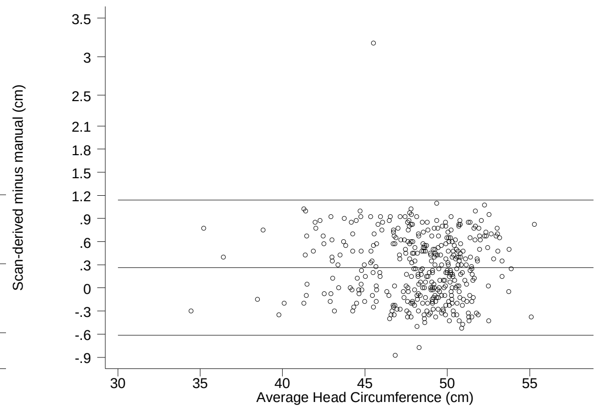
Pitman's Test: r = 0.188, n = 82, p = 0.091
Limits of agreement: -0.412 to 2.062
Mean difference: 0.825 (CI 0.689 to 0.961)

D.



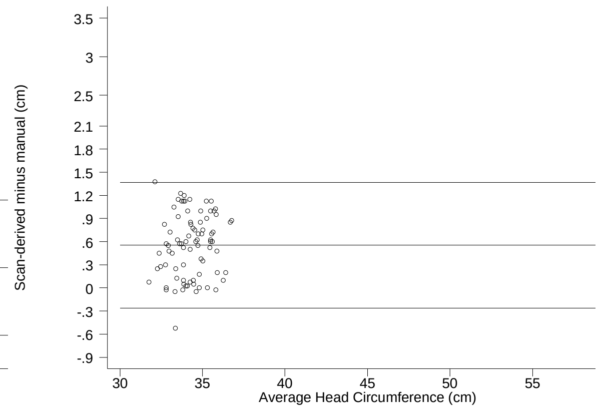
Pitman's Test: r = -0.232, n = 474, p = 0.000
Limits of agreement: -0.581 to 1.206
Mean difference: 0.312 (CI 0.272 to 0.353)

E.

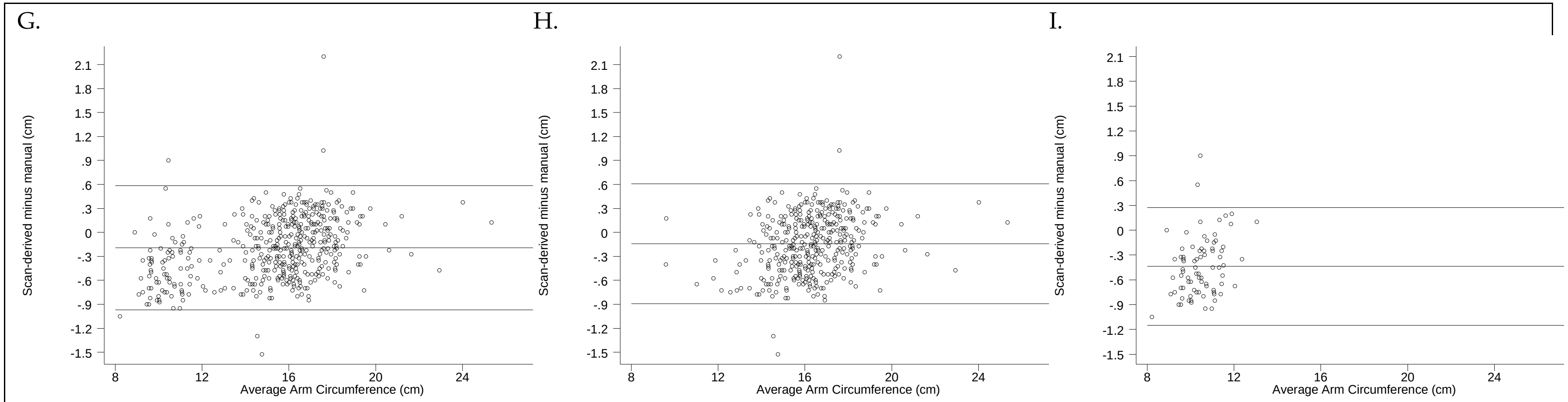


Pitman's Test: r = -0.044, n = 392, p = 0.386
Limits of agreement: -0.616 to 1.140
Mean difference: 0.262 (CI 0.218 to 0.306)

F.



Pitman's Test: r = 0.132, n = 82, p = 0.237
Limits of agreement: -0.261 to 1.367
Mean difference: 0.553 (CI 0.464 to 0.642)



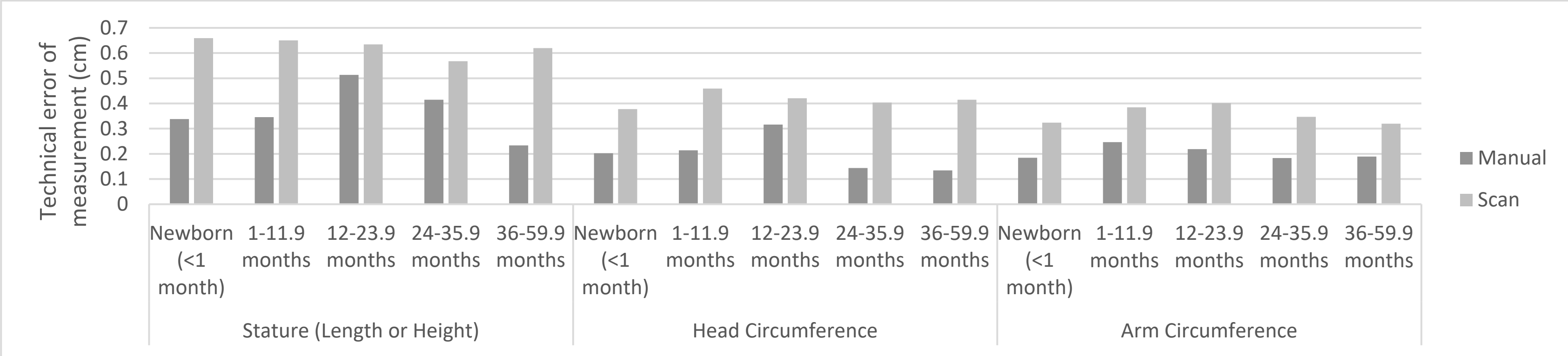
Pitman's Test: r = 0.378, n = 474, p = 0.000
Limits of agreement: -0.970 to 0.583
Mean difference: -0.193 (CI -0.228 to -0.158)

Pitman's Test: r = 0.259, n = 392, p = 0.000
Limits of agreement: -0.893 to 0.608
Mean difference: -0.142 (CI -0.180 to -0.105)

Pitman's Test: r = 0.291, n = 82, p = 0.008
Limits of agreement: -1.149 to 0.274
Mean difference: -0.437 (CI -0.516 to -0.359)

Figure 4-1 Bland-Altman plots of best-estimate manual measurements subtracted from single-scan measurements and related statistics, BINA 2017

A. Intra-observer



B. Inter-observer

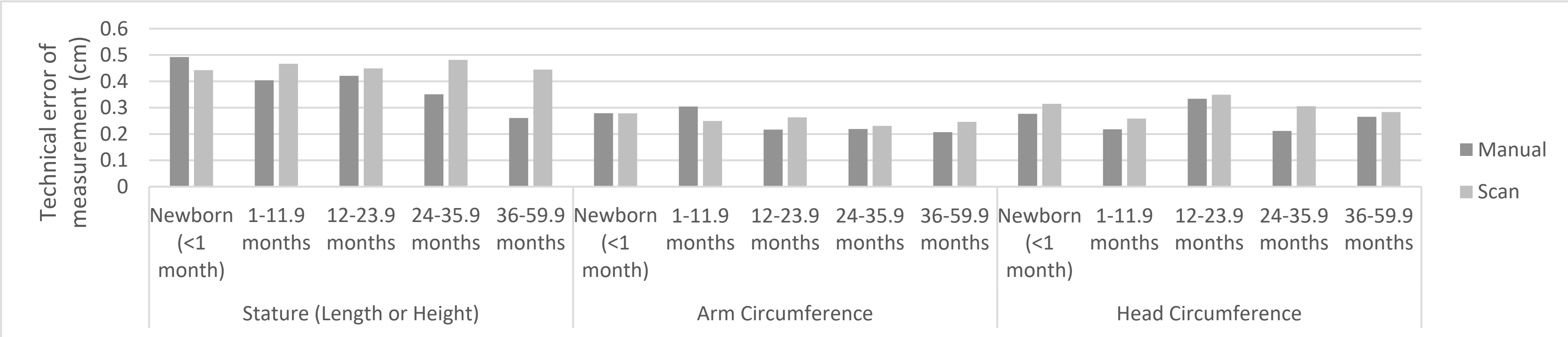


Figure 4-2 Intra- and inter-observer technical error of measurement (TEM) for scan-derived and manual measurements

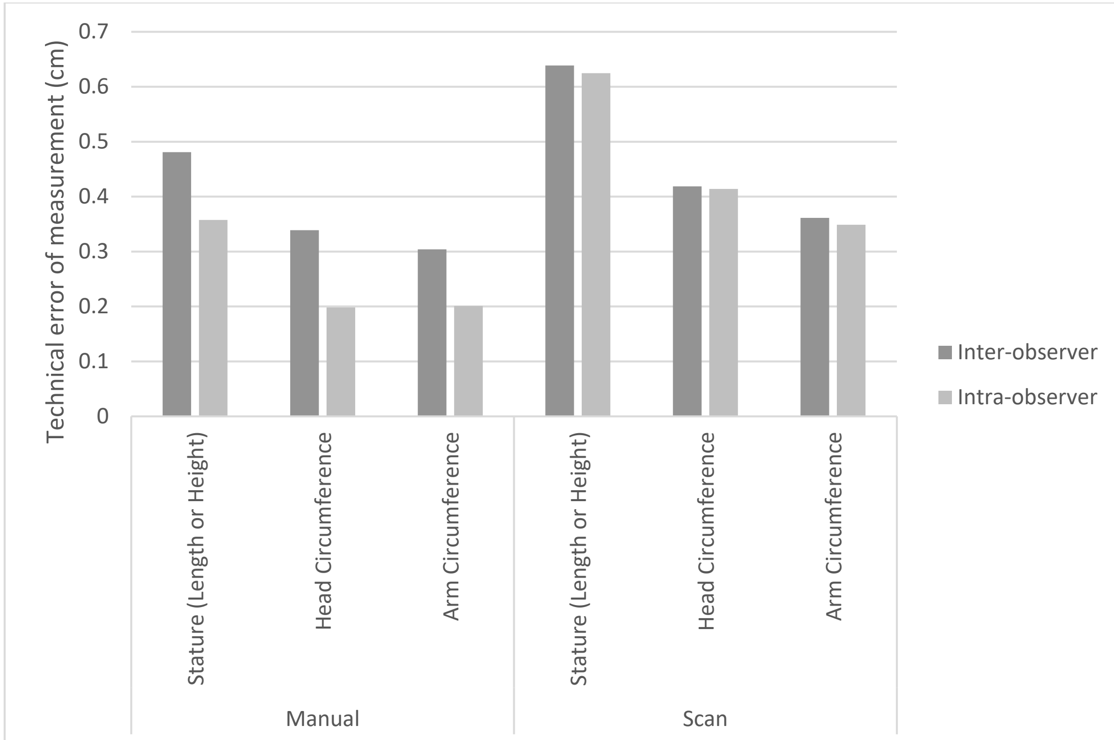


Figure 4-3 Single-measurement inter-observer technical error of measurement (TEM) versus intra-observer TEM for scan-derived and manual measurements

Table 4-2 Z-score mean, standard deviation (SD) and prevalence by selected z-score-for-age cutoffs among children 1-59.9 months of age

Row Labels	Sample Size	Z-score Mean	Z-score SD	Prevalence by z-score-for-age cutoff	
Stature				HAZ <-1 SD	HAZ <-2SD
Best-estimate Manual	392	-0.25	1.10	21.9	4.6
Single Manual	392	-0.24	1.10	22.7	4.3
Unadjusted Single Scan	392	-0.07	1.12	17.6	3.8
Adjusted Single Scan	392	-0.25	1.13	23.5	5.4
Adjusted Repeated Scan	392	-0.25	1.11	22.7	4.6
Head Circumference				HCZ >1 SD	HCZ >2 SD
Best-estimate Manual	392	0.34	1.02	27.3	3.3
Single Manual	392	0.34	1.04	27.3	3.8
Unadjusted Single Scan	392	0.53	1.07	32.9*	8.4***
Adjusted Single Scan	392	0.34	1.08	26.8	5.4
Adjusted Repeated Scan	392	0.35	1.03	26.5	3.8
Arm Circumference				ACZ >1 SD	ACZ >2 SD
Best-estimate Manual	385	0.78	0.94	41.6	9.6
Single Manual	385	0.78	0.97	41.8	9.9
Unadjusted Single Scan	385	0.67	1.04	37.9	9.1
Adjusted Single Scan	385	0.77	1.03	41.6	10.6
Adjusted Repeated Scan	385	0.76	1.01	40.5	11.4

*,***Significantly different from Best-estimate manual prevalence with Chi-Square at p<.10 and p<.01

Table 4-3 Sensitivity and specificity of adjusted, scan-derived measures when compared to best-estimate manual measures among children 1-59.9 months of age

	Sensitivity	Specificity	ROC Area (Average of sensitivity and specificity)	ROC Area 95% CI	
				Lower Limit	Upper Limit
Stature (HAZ < -1 SD)					
Single Manual	0.95	0.98	0.97	0.94	0.99
Single Scan	0.92	0.96	0.94	0.91	0.97
Repeated Scan	0.93	0.97	0.95	0.92	0.98
Head Circumference (HCZ >1 SD)					
Single Manual	0.94	0.98	0.96	0.94	0.99
Single Scan	0.84	0.95	0.89	0.86	0.93
Repeated Scan	0.87	0.96	0.92	0.88	0.95
Arm Circumference (ACZ>1 SD)					
Single Manual	0.93	0.95	0.94	0.91	0.96
Single Scan	0.91	0.94	0.93	0.90	0.95
Repeated Scan	0.93	0.96	0.95	0.92	0.97

Chapter 5 . Supplementary Materials for Accuracy and Reliability of a Low-Cost, Handheld 3D Imaging System for Child Anthropometry

5.1. Supplementary Methods

5.1.1. Accuracy

We assessed the accuracy of 3D imaging by comparing scan-derived measurements to the best-estimate from manual measurement. We calculated average bias with the following formula:

$$AverageBias = \Sigma_{i=1}^{N} (M_{i1} - M_{iF1})/N$$

, where M_{i1} is the scan-derived measurement and M_{iF1} is the best-estimate for each child. We used SPSS 20 (IBM Corp., Armonk, NY, USA) to test statistical significance of average bias with a two-sided, paired t-test with alpha of 0.05. Average bias is a metric of systematic bias. We also carried out Sign Tests — another metric of systematic bias that tests whether there were the same number of positive and negative differences using a Binomial Test.

Using StataSE 13's (StataCorp, College Station, TX, USA) *baplot* module we created Bland-Altman (BA) Plots (25) to assess if accuracy remained constant across different child body sizes and to look at random bias. For the y-axis of the BA Plot we subtracted the best-estimate from the single-scan value, and for the x-axis we used the mean of single-scan and best-estimate. We used Pitman's Test of Difference in Variance (33) to test the correlation between accuracy and the size of the child, and we calculated and plotted Limits of Agreement, which is the 95% precision interval for individual differences and is a metric of random bias. We disaggregated analysis based on age groups corresponding to animator's models (<1 month and 1-59.9 months). If accuracy was not consistent across different sizes, indicated by a statistically significant Pitman's Test, we carried out the additional step of regressing the difference on the second single-scan as suggested by Bartlett and Frost (33); they advocated for the use of linear regression to rule out difference in SD as the cause of a statistically significant Pitman's Test, proving that accuracy is truly biased.

We tested for differences in accuracy by caregiver-reported race and observed hairstyle using One Way Analysis of Variance (ANOVA). For comparing accuracy between races and hairstyles we used all-scan and best-estimate measurements to reduce variance and improve power. We coded hairstyle as a binary variable based on visual inspection of scans to identify children with protruding hair or hairstyles that could potentially affect scan-derived measurements. For race we defined three categories, namely Black, White, and Other; with the latter including Asian, Other, Multiple, and Not Reported. For ANOVA we tested the assumption of homogeneity of variances with Levene's Test (34), and used Welch Test when variances were not homogenous (35). We did not further disaggregate race by ethnicity because of sample size constraints.

5.1.2. Reliability

We calculated intra-observer technical error of measurement (TEM) for scan-derived and manual measurements with the following formula:

$$TEMintra = \sqrt{\sum_{i=1}^{N}(M_{i1} - M_{i2})^2 / 2 * N}$$

, where N was the number of children and $M_{i1} M_{i2}$ were the first and second measurements for one child (repeated measurements) by one observer. For inter-observer TEM we compared average measurements with the following formula:

$$TEMinter = \sqrt{\sum_{i=1}^{N}\left({}^{M_{ij1} + M_{ij2}}/_{2} - {}^{M_{ik1} + M_{ik2}}/_{2}\right)^2 / 2 * N}$$

, where $M_{ij1}\ M_{ij2}$ were repeated measurements from one observer and $M_{ik1} M_{ik2}$ were repeated measurements from another observer of the same child. We also analyzed inter-observer TEM based on single measurements. For inter- and intra-observer TEM we calculated relative TEM, or %TEM, by dividing TEM by the measurement mean and converting to a percent; and used SPSS 20 to calculate the Intraclass Correlation Coefficient based on absolute agreement. Total TEM combined inter- and intra-observer TEM with the following formula:

$$TotalTEM = \sqrt{TEMintra^2 + TEMinter^2}$$

, where $TEMinter$ came from repeated measurements. We calculated an overall R, or Coefficient of Reliability, with the following formula from Ulijaszek (26):

$$R = 1 - \frac{TotalTEM^2}{SD^2}$$

, where SD^2 was the pooled variance of the four measurements used to calculate Total TEM. We divided the sample into five age groups considering sample size for disaggregated analysis of reliability.

5.1.3. Z-scores and Classification

We calculated z-scores based on the 2006 WHO CGS using WHO SPSS 20 macros (36). Anthropometric indices and nutritional status included in sample characteristics were based on best-estimate manual measurements. For additional analysis of z-scores and classification, we selected children 1-59.9 months of age to evaluate the animator model with sufficient sample size. To compare z-score means, standard deviations and prevalence we calculated z-scores from the best-estimate, single-manual, single-scan and repeated-scan. We tested for statistical significance of prevalence differences using Chi-Square and comparing to the prevalence from the best-estimate. We included both adjusted and unadjusted scan derived measurements, with adjustment made by subtracting or adding the average bias found in our analysis of accuracy from/to each observation. We evaluated individual level classification with Stata's *diagt* module, reporting sensitivity and specificity for stature-, head circumference-, and arm circumference-for-age below or above 1 standard deviation. For individual level classification we considered the best-estimate "true" nutritional status. We selected z-score SD cutoffs based on having an adequate percentage of the sample selected by the cutoff for meaningful comparisons of z-scores and classification. We did not have sufficient sample size to analyze differences in z-scores and classification for newborns, or to analyze individual level classification at lower cutoffs, such as <-2 SD.

5.2 Supplementary Discussion

5.2.1. Single- vs Repeated-Scan Reliability

We designed the 3D imaging system to derive measurements based on a single scan session. In this study we found that single-scan was less reliable than single manual measurements, but that averaging two measurements from two scan sessions (repeated-scan) improved TEM and achieved reliability comparable to repeated manual measurements. Repeated-scan achieved excellent reliability and single-scan was less reliable, but the difference between single and repeated scan reliability was not large (indicated by difference between inter- and intra-observer TEMs in table S3). Furthermore, the differences in z-score SDs and percentages below or above SD cutoffs for single- and repeated-scan were small and statistically insignificant. Additional research that is adequately powered to detect small differences in prevalence is needed to determine if the protocol for deriving measurements from 3D imaging should be changed from using a single-scan to repeated-scan. We believe the use of repeated scans is feasible and would not be overly burdensome for the anthropometrist or child because repeating scans would only add ~one minute to measurement time. We will be able to quantify the effect of a change in protocol on the time required for collecting child anthropometry in a time-motion study that is underway to compare 3D imaging with manual measurements.

We did not have maximum allowable difference triggers to improve reliability of 3D imaging because scan-derived measurements are not immediately available, and so did not use triggered, third manual measurements for calculating inter-observer TEM in this study. However, we compared inter-observer TEM results from this study to a previous publication where we did use the third manual measurements (23) and we found that triggers did little to improve our manual measurement reliability; the reliability of repeated-scan is similar to manual measurements with or without the use of maximum allowable difference triggers for manual measurements.

5.2.2. The Effect of Reliability and Accuracy on Monitoring and Classification

We showed that 3D imaging produced reliable measurements; we also need to consider how this reliability would affect the quality of anthropometric data outside of a research setting. Most large-scale surveys do not take repeated measurements and so cannot be analyzed for reliability, but reliability is directly related to z-score SD. Reliability is a metric of random error, and as random error increases z-score SD increases. If z-score SD is too high, there may be overdispersion, which can cause overestimation of prevalence below or above z-score cutoffs. High quality surveys have z-score SD between 0.9-1.1 (37, 38); in this study 3D imaging z-score SD was within 0.9-1.1 for all measures. An evaluation of the quality of 52 DHS found that 1 out of 52 had HAZ SD between 0.9-

1.1, and 30 out of 52 had HAZ SD above 1.5 (17). A SD of 1.5 can lead to overestimation of prevalence. A recent study found that DHS and MICS carried out in Western and Central Africa from 1990-2012 may have overestimated the prevalence of stunting (HAZ <-2 SD) by ~10 percentage points on average (38). Overdispersion from poor quality anthropometry can also result in overestimation of overweight and obesity. Poor quality anthropometric data is common in large-scale surveys, and quality is variable between countries and between surveys in the same country; making it difficult to meaningfully compare countries or analyze trends over time. The reliability of 3D imaging and manual anthropometry in BINA was good enough to make meaningful comparisons between countries and over time. Our reliability findings also indicate that 3D imaging can be used for growth monitoring. Considering a well check schedule of visits at 1, 2, 4, 6, 9, 12, 15, 18 and 24 months; the differences in stature between visits is 3.2-5.5 cm according to the 2006 WHO CGS median. The inter-observer TEM for repeated-scan indicates that repeated measurements from 3D imaging can be within 1.5 cm 99.9% of the time — 3D imaging random error would not cause overlapping estimates in growth monitoring.

We measured accuracy with average bias, a metric of systematic bias. In the MGRS average bias was considered acceptable if it was within ±2.8 times the expert intra-observer TEM (27); based on MGRS criteria 3D imaging accuracy was acceptable for all measures in our study. However, our findings indicate that the criteria used to assess average bias for the 2006 WHO CGS may be too lenient for the purposes of our study. For head circumference the acceptable average bias was set at ±0.34 cm in the MGRS (27). In our study the average bias for head circumference from scan-derived measurements among children 1-59.9 months of age was well within the acceptable range at 0.26, but our HC average bias led to a statistically significant five percentage point overestimation of the percent of the population above 1 SD and 2 SD when compared to manual measurement (Table 2). Since 3D imaging produced measurements that were essentially as reliable as manual, we can conclude that the differences in prevalence are due to inaccuracy, showing that HC average bias of 0.26 is likely not acceptable. The acceptability of average bias should be based on producing similar prevalence estimates and avoiding misclassification of individuals. For prevalence estimates there is no single criteria for how close is close enough, but we can consider how prevalence is used to posit reasonable criteria. The WHO cutoff for a moderate public health problem for wasting is 5%, which is ~2.5 percentage points from what is expected in a healthy population. It is reasonable to say that differences in prevalence between an index test and a reference standard cannot exceed 2.5 percentage points for child anthropometry. From another perspective, differences between the two methods should ideally not be statistically detectable in a common, large scale-survey; and should allow for meaningful program evaluation, which would put the level closer to ~1 percentage point. A typical program goal may be to reduce stunting by one to two percentage points annually over three to five years. A program effect may be statistically masked by error over two percentage points. After adjustment prevalence differences between repeated-scan and best-estimate in our study were less than one percentage point for stature and head circumference, and less than two percentage points for MUAC; which is probably good enough for common uses of prevalence estimates. For individual classification sensitivity and specificity above 0.90 is generally considered excellent. We were able to achieve excellent classification at the individual level, but only after adjusting measurements to remove systematic bias.

5.2.3. Bias Hypotheses

3D imaging overestimated head circumference, while underestimating MUAC. Consideration of the protocol for manual measurements provides a hypothesis for the different directions of average bias. When measuring head circumference, anthropometrists were instructed to apply enough tension to compress the hair and underlying soft tissue. For MUAC, the measuring tape was s meant to be flush with the skin without any compression. 3D imaging may have overestimated head circumference because compression of soft tissue was not taken into account during calibration. If compression is the source of bias for head circumference, we may expect no average bias for MUAC because there is no compression for MUAC. Interestingly, for MUAC ≥ 16.7 cm we found essentially no systematic bias, and it was only among smaller children (MUAC 9.6-15.1 cm) that we found substantial underestimation. It is possible that underestimation by 3D imaging for MUAC was caused by manual measurement error. It can be more difficult to measure MUAC on younger children because increased adiposity along with decreased cooperation make it more difficult to maintain the measuring tape flush against the skin. It is possible that BINA anthropometrists systematically left small gaps between the skin and measuring tape when measuring MUAC of younger children, and that manual measurements are less accurate than 3D imaging for young children.

The protocol for manually measuring head circumference may also help to explain another finding: 3D imaging was better at selecting the same children above or below SD cutoffs for stature and MUAC than it was for HC. The protocol for measuring manual HC calls for shifting the measuring tape up and down to select the largest circumference; there is no fixed position. The current 3D imaging system is based on measuring fixed, pre-specified points; and does not account for being able to shift the measuring tape up and down. To achieve excellent sensitivity 3D imaging system software may have to be developed to identify the largest circumference and not just rely on pre-specified points.

When processing scans we returned the final 3D model to a "neutral" position to take stature measurements. Comparing the "neutral" position to WHO measuring protocol, we did not take into account the Frankfort Plane and the back of the head was not aligned with the heels, buttocks, and back of shoulders. It is not clear what effect model positioning had on the accuracy of scan-derived stature and if positioning is a reasonable hypothesis for overestimation. Another possibility for overestimation is that 3D imaging did not adequately take into account hair compression that occurs when taking manual stature measurements.

5.2.4. Study Limitations – Automation and Anthropometrists

Per study protocol children were required to adopt fixed positions when taking scans, but anthropometrists could not touch the child because touching impaired the ability to process the scans into measurements. During pretesting we discovered that an anthropometrist could help to put the child into position if both the measurer and the child held a common object, such as a large, plastic spoon. However, for some children this approach did not work due to lack of cooperation, thus we further altered the protocol by allowing anthropometrists to take excess scans. When anthropometrists took extra scans, they then selected the 12 best scans for each child and deleted the rest. Another departure from the original protocol was that anthropometrists could not always take scans in the specified sequence of orientation: front and then back. With many children anthropometrists had to be opportunistic and take scans of a child when they were still, ignoring their orientation. For processing scans we had to manually code orientation after data collection. Processing scans was initially designed to be fully automated; the manual input to delete excess scans and define the child's orientation make our findings less generalizable because there is more potential for human error, and we do not know if the number of scans taken per child affected accuracy or reliability. Further development of the 3D imaging system scanning and processing software is needed to achieve full automation.

Our primary interest in researching 3D imaging for child anthropometry was to improve the quality of anthropometric data. Anthropometrists in BINA, who were well educated, highly motivated, and well-trained, achieved high quality anthropometric data with both 3D imaging and manual measurement. Compared to manual measurement, we spent substantially less time on training and supervision for 3D scanning, but it is not yet possible to determine if 3D imaging would lead to better quality in a setting that produces poor quality anthropometric data from manual measurement. Qualitative research on

BINA anthropometrists' experiences using 3D scanners is currently underway and this may help to provide some evidence on the potential of 3D imaging to improve anthropometric data quality. However, conclusive evidence may not be available until the use of 3D imaging for infant and young child anthropometry is evaluated in a real-world application, such as a large-scale survey.

5.3 Supplementary Figures and Tables

Figure 5-1. 3D scan arm poses for children two years of age and over, BINA 2017

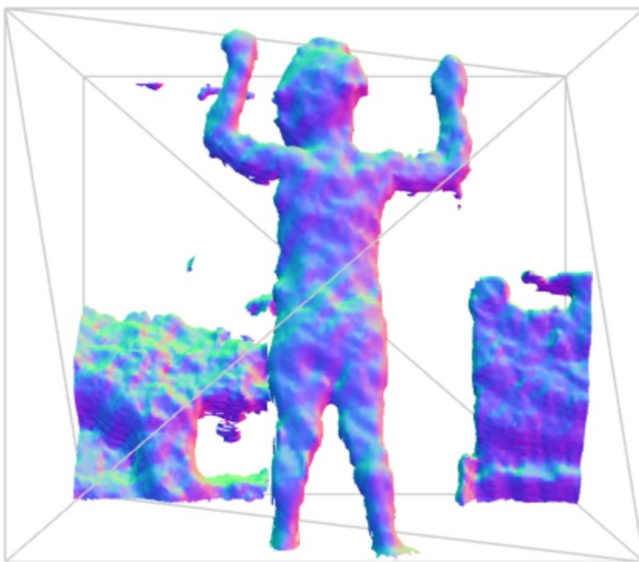


Figure 5-2. 3D scan as it appears before processing, BINA 2017

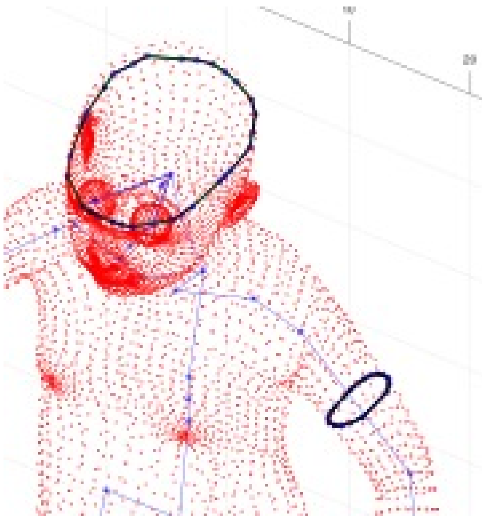


Figure 5-3. S3 Points (in black) selected on base model to measure head and arm circumference

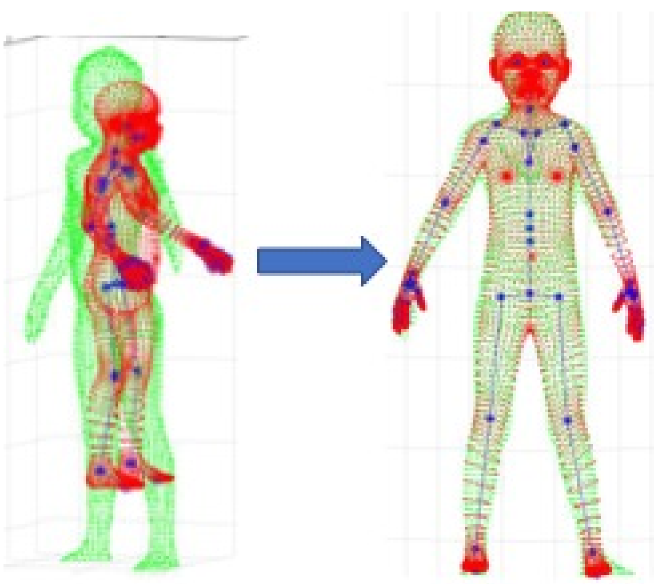


Figure 5-4. The basic fitting process[a]

[a]Scan data is in green, articulated model surface in red, "bones" and "joints" in blue. On the left, the initial size and pose of model relative to data. On the right, the model has been automatically sized and posed to fit the scan data.

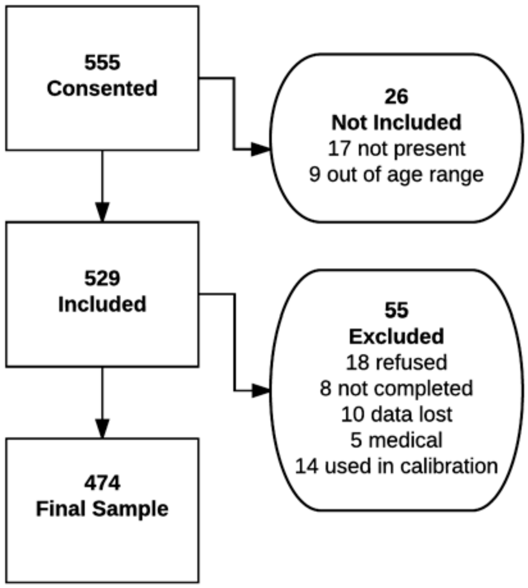


Figure 5-5. Flow of participants, BINA 2017

Table 5-1. Accuracy of scan-derived measurements when compared to best-estimate, manual measurements among all children under five years of age, BINA 2017

	Paired T-Test					Percent of Positive Differences[b]
Measurer and Observation	**Mean from Scan**	**Mean Difference (Scan-Manual)**	**T value for Difference From 0[a]**	**Difference 95% Limits of Agreement**		
				Lower Limit	**Upper Limit**	
Stature (Length or Height)						
Single scan session						
Measurer 1 (M1) Observation 1 (O1)	82.9	0.61	20.2	-0.7	1.9	78%
M1, Observation 2 (O2)	82.9	0.56	19.2	-0.7	1.8	78%
Measurer 2 (M2), O1	82.9	0.58	20.1	-0.7	1.8	80%
M2, O2	82.9	0.60	21.1	-0.6	1.8	80%
Average from two scan sessions						
M1	82.9	0.59	26.8	-0.4	1.5	90%
M2	82.9	0.59	29.2	-0.3	1.5	91%
Average from four scan sessions						
M1&M2						
M1&M2	82.9	0.59	39.2	-0.1	1.2	97%
Head Circumference						
Single scan session						
M1, O1	46.1	0.31	15.2	-0.6	1.2	72%
M1, O2	46.1	0.32	16.9	-0.5	1.2	73%
M2, O1	46.1	0.32	16.7	-0.5	1.1	77%
M2, O2	46.1	0.34	16.4	-0.6	1.2	73%
Average from two scan sessions						
M1	46.1	0.32	22.1	-0.3	0.9	83%
M2	46.1	0.33	22.1	-0.3	1.0	84%
Average from four scan sessions						
M1 & M2	46.1	0.32	29.9	-0.1	0.8	93%
Arm Circumference						
Single scan session						
M1, O1	15.2	-0.19	-10.9	-1.0	0.6	34%
M1, O2	15.2	-0.20	-12.4	-0.9	0.5	35%
M2, O1	15.2	-0.20	-11.7	-0.9	0.5	33%
M2, O2	15.2	-0.17	-10.1	-0.9	0.6	36%
Average from two scan sessions						
M1	15.2	-0.20	-15.6	-0.7	0.4	25%
M2	15.2	-0.19	-14.5	-0.7	0.4	28%
Average from four scan sessions						

M1 & M2	15.2	-0.19	-19.7	-0.6	0.2	20%

[a]All mean differences significantly different from zero at p<.0001
[b]From binomial test percentages are all significantly different from 50% at p<.0001

Table 5-2. Accuracy by race and hairstyle considering best-estimate manual measurements and scan-derived measurements from all sessions among children 1 to 59.9 months of age, BINA 2017

Race or Hairstyle	**n**	**Mean Difference** (scan-manual)	**Standard Deviation (SD)**	**One-Way Analysis of Variance**	
				F	Significance
Stature by Race				0.30	0.74
Black	136	0.53	0.33		
White	126	0.55	0.32		
Other	130	0.56	0.31		
Head Circumference by Race				1.37	0.25
Black	136	0.30	0.24		
White	126	0.26	0.17		
Other	130	0.27	0.22		
Head Circumference by Hairstyle				0.39	0.53
Large hair	347	0.28	0.22		
Not large hair	45	0.26	0.21		
Arm Circumference by Race				0.02	0.98
Black	136	-0.15	0.20		
White	126	-0.15	0.18		
Other	130	-0.14	0.19		

Asymptomatically F distributed

Table 5-3 Intra-observer reliability and inter-observer reliability based on repeated manual measurements and repeated scan sessions by age group, BINA 2017

Intra-observer Reliability											
	Sample Size	**Average (cm)**		**Mean Absolute Difference (cm)**		**Technical Error of Measurement (TEM) (cm)**		**Relative TEM (%TEM)**		**Intraclass Correlation Coefficient (ICC)**	
Row Labels		**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**
Stature (Length or Height)											
All (0-4.9 years)	948	82.3	82.9	0.3	0.7	0.36	0.62	0.4	0.8	1.00	1.00
Newborn (<1 month)	164	48.8	49.6	0.3	0.8	0.34	0.66	0.7	1.3	0.96	0.86
1-11.9 months	132	66.2	66.8	0.4	0.8	0.35	0.65	0.5	1.0	1.00	0.99
12-23.9 months	150	81.2	81.8	0.4	0.7	0.51	0.63	0.6	0.8	0.99	0.98
24-35.9 months	170	90.3	90.8	0.3	0.7	0.41	0.57	0.5	0.6	0.99	0.98
36-59.9 months	332	101.7	102.2	0.2	0.7	0.23	0.62	0.2	0.6	1.00	0.99
Head Circumference											
All (0-4.9 years)	948	45.7	46.1	0.2	0.5	0.20	0.41	0.4	0.9	1.00	1.00
Newborn (<1 month)	164	34.0	34.6	0.2	0.4	0.20	0.38	0.6	1.1	0.97	0.89
1-11.9 months	132	43.1	43.5	0.2	0.5	0.21	0.46	0.5	1.1	0.99	0.97
12-23.9 months	150	47.5	47.8	0.2	0.5	0.32	0.42	0.7	0.9	0.96	0.94
24-35.9 months	170	48.8	49.0	0.1	0.5	0.14	0.40	0.3	0.8	0.99	0.94
36-59.9 months	332	50.2	50.5	0.1	0.5	0.13	0.41	0.3	0.8	0.99	0.93
Arm Circumference											
All (0-4.9 years)	948	15.4	15.2	0.2	0.4	0.20	0.35	1.3	2.3	0.99	0.99
Newborn (<1 month)	164	10.7	10.3	0.2	0.4	0.18	0.32	1.7	3.2	0.95	0.88
1-11.9 months	132	14.7	14.4	0.2	0.4	0.25	0.38	1.7	2.7	0.98	0.95
12-23.9 months	150	15.9	15.8	0.2	0.5	0.22	0.40	1.4	2.5	0.97	0.91

24-35.9 months	170	16.6	16.5	0.2	0.0	0.18	0.35	1.1	2.1	0.98	0.93
36-59.9 months	332	17.2	17.1	0.2	0.3	0.19	0.32	1.1	1.9	0.99	0.96

Inter-observer Reliability (average of repeated measures)

	Sample Size	**Average in cm**		**Mean Absolute Difference (cm)**		**Technical Error of Measurement (TEM)**		**Relative TEM (%TEM)**		**Intraclass Correlation Coefficient (ICC)**	
Row Labels	**manual**	**manual**	**scan**	**manual**	**scan**	**manual**	**Scan**	**manual**	**scan**	**manual**	**scan**
Stature (Length or Height)											
All (0-4.9 years)	474	82.3	82.9	0.4	0.5	0.37	0.46	0.5	0.5	1.00	1.00
Newborn (<1 month)	82	48.8	49.6	0.5	0.5	0.49	0.44	1.0	0.9	0.92	0.93
1-11.9 months	66	66.2	66.8	0.4	0.5	0.40	0.47	0.6	0.7	1.00	0.99
12-23.9 months	75	81.2	81.8	0.4	0.5	0.42	0.45	0.5	0.5	0.99	0.99
24-35.9 months	85	90.3	90.8	0.3	0.6	0.35	0.48	0.4	0.5	0.99	0.99
36-59.9 months	166	101.7	102.2	0.3	0.5	0.26	0.44	0.3	0.4	1.00	0.99
Head Circumference											
All (0-4.9 years)	474	45.7	46.1	0.3	0.4	0.26	0.30	0.6	0.7	1.00	1.00
Newborn (<1 month)	82	34.0	34.6	0.3	0.4	0.28	0.31	0.8	0.9	0.94	0.92
1-11.9 months	66	43.1	43.5	0.2	0.3	0.22	0.26	0.5	0.6	0.99	0.99
12-23.9 months	75	47.5	47.8	0.3	0.4	0.33	0.35	0.7	0.7	0.96	0.95
24-35.9 months	85	48.8	49.0	0.2	0.4	0.21	0.31	0.4	0.6	0.98	0.96
36-59.9 months	166	50.2	50.5	0.2	0.3	0.27	0.28	0.5	0.6	0.97	0.97
Arm Circumference											
All (0-4.9 years)	474	15.4	15.2	0.3	0.3	0.24	0.25	1.6	1.7	0.99	0.99
Newborn (<1 month)	82	10.7	10.3	0.3	0.3	0.28	0.28	2.6	2.7	0.89	0.90
1-11.9 months	66	14.7	14.4	0.3	0.3	0.30	0.25	2.1	1.7	0.96	0.98
12-23.9 months	75	15.9	15.8	0.2	0.3	0.22	0.26	1.4	1.7	0.97	0.96
24-35.9 months	85	16.6	16.5	0.2	0.3	0.22	0.23	1.3	1.4	0.97	0.97
.9 months	166	17.2	17.1	0.2	0.3	0.21	0.25	1.2	1.4	0.98	0.98

References (Chapters 4 and 5)

1. World Health Organization., UNICEF. WHO child growth standards and the identification of severe acute malnutrition in infants and children : joint statement by the World Health Organization and the United Nations Children's Fund. Geneva: World Health Organization; 2009. 11 p. p.

2. NHS Commissioning Board. Clinical Commissioning Policy: Complex and Specialised Obesity Surgery. London: NHS Commissioning Board; 2013.

3. Mehta NM, Corkins MR, Lyman B, Malone A, Goday PS, Carney LN, et al. Defining pediatric malnutrition: a paradigm shift toward etiology-related definitions. JPEN J Parenter Enteral Nutr. 2013;37(4):460-81.

4. Patel V, Romano M, Corkins MR, DiMaria-Ghalili RA, Earthman C, Malone A, et al. Nutrition Screening and Assessment in Hospitalized Patients: A Survey of Current Practice in the United States. Nutr Clin Pract. 2014;29(4):483-90.

5. Voss LD, Mulligan J, Betts PR, Wilkin TJ. Poor growth in school entrants as an index of organic disease: the Wessex growth study. BMJ. 1992;305(6866):1400-2.

6. Gibson RS. Principles of nutritional assessment. 2nd ed. New York: Oxford University Press; 2005. xx, 908 p. p.

7. Wit JM, Himes JH, van Buuren S, Denno DM, Suchdev PS. Practical Application of Linear Growth Measurements in Clinical Research in Low- and Middle-Income Countries. Horm Res Paediatr. 2017.

8. Setboonsarng S. Child Malnutrition as a Poverty Indicator: An Evaluation in the Context of Different Development Interventions in Indonesia. ADB Institute Discussion Paper 21. Tokyo: Asian Development Bank Institute; 2005.

9. Bleakley H. Health, Human Capital, and Development. Annu Rev Econom. 2010;2:283-310.

10. World Bank. World Development Indicators 2016. Washington, D.C.: World Bank; 2016.

11. IEG-World Bank. The World Bank's Country Policy and Institutional Assessment: An Evaluation. Washington, D.C.: World Bank; 2010.

12. Sebo P, Haller D, Pechere-Bertschi A, Bovier P, Herrmann F. Accuracy of doctors' anthropometric measurements in general practice. Swiss Med Wkly. 2015;145:w14115.

13. Sebo P, Beer-Borst S, Haller DM, Bovier PA. Reliability of doctors' anthropometric measurements to detect obesity. Prev Med. 2008;47(4):389-93.

14. Yin H, Dai Y, Li H, Xie X, Ren H. The test–re-test reliability of routine infant anthropometry at primary care hospitals in Chongqing, PR China. Annals of Human Biology. 2013;40(4):309-17.

15. Lipman TH, Hench KD, Benyi T, Delaune J, Gilluly KA, Johnson L, et al. A multicentre randomised controlled trial of an intervention to improve the accuracy of linear growth measurement. Arch Dis Child. 2004;89(4):342-6.

16. Gerner B, McCallum Z, Sheehan J, Harris C, Wake M. Are general practitioners equipped to detect child overweight/obesity? Survey and audit. J Paediatr Child Health. 2006;42(4):206-11.

17. Assaf S, Kothari, M., Pullum, T. An Assessment of the Quality of DHS Anthropometric Data, 2005-2014. DHS Methodological Reports No 16. Rockville, MD, USA: ICF International; 2015.

18. Pullum TW. An Assessment of the Quality of Data on Health and Nutrition in the DHS Surveys, 1993-2003. Methodological Reports No 6. Calverton, MD, USA: Macro International Inc.; 2008.

19. Corsi DJ, Perkins JM, Subramanian SV. Child anthropometry data quality from Demographic and Health Surveys, Multiple Indicator Cluster Surveys, and National Nutrition Surveys in the West Central Africa region: are we comparing apples and oranges? Glob Health Action. 2017;10(1):1328185.

20. WHO Expert Committee on Physical Status. The Use and Interpretation of Anthropometry. WHO Technical Report Series 854. Geneva: World Health Organization; 1995.

21. USAID. Anthropometric Data in Population-Based Surveys, Meeting Report, July 14-15, 2015. Washington, D.C., USA: FHI 360/FANTA; 2016.

22. de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr Bull. 2004;25(1 Suppl):S27-36.

23. Conkle J, Ramakrishnan, U., Suchdev, P., Flores-Ayala, R., Martorell, R. Methods and quality of manual anthropometry in the 3D Body Imaging for Nutritional Assessment Study (BINA). Manuscript in preparation. 2017.

24. De Onis M, World Health Organization. Dept. of Nutrition for Health and Development. WHO child growth standards : length/height-for-age, weight-for-age, weight-for-length, weight -for-height and body mass index-for-age : methods and development. Geneva: World Health Organization; 2006.

25. Martin Bland J, Altman D. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet. 1986;327(8476):307-10.

26. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. The British journal of nutrition. 1999;82(3):165-77.

27. WHO Multicentre Growth Reference Study Group. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl. 2006;450:38-46.

28. Jones PR, West GM, Harris DH, Read JB. The Loughborough anthropometric shadow scanner (LASS). Endeavour. 1989;13(4):162-8.

29. Barnes R, editor The Body Volume Index (BVI): Using 3D Scanners to Measure and Predict Obesity. International Conference on 3D Body Scanning Technologies; 2010; Lugano, Switzerland: Hometrica Consulting.

30. Weathers WM, Khechoyan D, Wolfswinkel EM, Mohan K, Nagy A, Bollo RJ, et al. A novel quantitative method for evaluating surgical outcomes in craniosynostosis: pilot analysis for metopic synostosis. Craniomaxillofac Trauma Reconstr. 2014;7(1):1-8.

31. 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;10(3):e0119430.

32. Soileau L, Bautista D, Johnson C, Gao C, Zhang K, Li X, et al. Automated anthropometric phenotyping with novel Kinect-based three-dimensional imaging method: comparison with a reference laser imaging system. Eur J Clin Nutr. 2016;70(4):475-81.

33. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol. 2008;31(4):466-75.

34. Brown MB, Forsythe AB. Robust Tests for the Equality of Variances. Journal of the American Statistical Association. 1974;69(346):364-7.

35. Welch BL. On the Comparison of Several Mean Values: An Alternative Approach. Biometrika. 1951;38(3/4):330-6.

36. WHO. WHO Anthro and macros Geneva: WHO; 2015 [Available from: http://www.who.int/childgrowth/software/en/.

37. SMART Action Against Hunger - Canada and Technical Advisory Group. The SMART Plausibility Check for Anthropometry. Toronto; 2015.

38. Grellety E, Golden MH. The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition. PloS one. 2016;11(12):e0168585.

Chapter 6 . A collaborative, mixed-methods evaluation of a low-cost, handheld 3D imaging system for child anthropometry

Joel Conkle [1,*], Kate Keirsey[2], Ashton Hughes[2], Jennifer Breiman[2], Usha Ramakrishnan [1, 2], Parminder S. Suchdev [1,2,3,4], and Reynaldo Martorell [1,2]

[1] Doctoral Program in Nutrition and Health Sciences. Laney Graduate School. Emory University. Atlanta, GA, USA; joel.conkle@emory.edu

[2] Hubert Department of Global Health. Rollins School of Public Health. Emory University. Atlanta, GA, USA

[3] Division of Nutrition, Physical Activity and Obesity. National Center for Chronic Disease Prevention and Health Promotion. U.S. Centers for Disease Control and Prevention. Atlanta, GA, USA

[4] Department of Pediatrics, School of Medicine, Emory University. Atlanta, GA, USA

* Correspondence: joelconkle@gmail.com; Tel.: +1-404-725-7138

Abstract:

3D imaging for body measurements (e.g., anthropometry) is regularly used for design of garments and ergonomic products. In recent years the development of low-cost 3D scanners provided an opportunity to extend the use of 3D imaging to the health sector. We developed and tested the AutoAnthro System, the first mobile, full-body, 3D imaging system designed specifically for child anthropometry. In a previous publication we showed that the AutoAnthro System produced reliable scan-derived measurements of child length, height, head circumference and arm circumference. This study evaluated the efficiency, invasiveness, and user experience of the newly developed 3D imaging system.

We used a mixed-methods, collaborative approach that included a quantitative time-motion study and qualitative interviews of anthropometrists. The time-motion study employed continuous observation of manual measurement and scanning based on milestone timing, and we designed and analyzed the qualitative component based on grounded theory from a constructivist point of view. For the qualitative component we analyzed in-depth interviews and a focus group discussion with line-by-line coding, memoing and network mapping. We collected data in February 2017.

We observed measuring and scanning of 22 children, and completed five in-depth interviews and one focus group discussion with participation from all five anthropometrists. We found that for cooperative children, anthropometrists considered the use of 3D imaging an easy, 'streamlined experience,' but with uncooperative children or when experiencing 'software glitches,' anthropometrists reported that capturing a good quality scan was out of their control. The mean time to complete a full set of scans was 68 seconds (standard deviation (SD) 29), compared to 135 seconds (SD 22) for a set of manual measurements (stature, head circumference, and arm circumference). We observed that crying was more common during manual measurement, and anthropometrist interviews confirmed that 3D imaging was less stressful for children than manual measurement.

Overall, the anthropometrists were not yet ready to completely abandon traditional, manual equipment for 3D scanners. Revising the AutoAnthro System to address anthropometrists' concerns on capturing good quality scans of uncooperative children should help to facilitate widespread use of 3D imaging for child anthropometry in the health sector.

6.1. Introduction

3D imaging for anthropometry was developed in 1989 for use in the garment industry, which relied on sizing surveys for pattern development (1). By the late 1990s a large-scale sizing survey provided scan-derived anthropometric data to manufacturers for design of garments and ergonomic products (2). 3D scanners are now commonplace in national sizing surveys, with multiple countries adopting the technology (3-5). 3D imaging has also been used for anthropometry in the health sector. Over the last decade multiple studies tested various 3D scanners for measurements of clinical interest, such as adult height (6), waist/hip circumference (7, 8), body fat (9, 10), body surface and volume (3, 11, 12), and body shape (13). Research on metabolic syndrome used scan-derived anthropometry (8, 14), and over the last few years large-scale epidemiological studies of adults used 3D scanners (13, 15). However, previous research on 3D imaging for anthropometry used expensive, stationary scanners, 3D imaging is not yet used in health and nutrition surveys, and its use in health facilities is limited to research and specialized purposes, such as cranial remolding orthoses (16).

The development of "light-coding" technology reduced the cost and size of 3D scanners and led to use in the gaming industry; and in 2013 a Kickstarter campaign funded the development of Structure Sensor (Occipital, San Francisco, CA, USA), an open-source 3D scanner that attaches to a tablet or phone. The development of low-cost, mobile, open-source scanners provided an opportunity to extend the use of 3D imaging to common uses of anthropometry in the health sector, such as nutritional screening and surveillance. We tested the AutoAnthro System, a tablet-based 3D imaging system for anthropometry designed for children under five years of age. In a previous publication we showed that the AutoAnthro System produced reliable, scan-derived measurements of child length, height, head circumference and arm circumference (17). This study evaluated the efficiency, invasiveness, and user experience of the newly developed 3D imaging system. The purposes of our research were to inform further development of AutoAnthro and to assess the potential for widespread use of 3D imaging for child anthropometry.

6.2. Materials and Methods

The Body Imaging for Nutritional Assessment Study (BINA) compared traditional, manual anthropometry to 3D imaging; and was approved by the Emory Institutional Review Board. For BINA we used AutoAnthro (BST, Atlanta, GA, USA), a custom software developed by Body Surface Translations (BST) for capturing and processing scans using the Structure Sensor. We calibrated AutoAnthro based on a sample of 36 children. We then carried out a cross-sectional, validation study on 474 apparently healthy children under five years of age. Five trained anthropometrists conducted all measurements and scans in daycares and medical facilities in Atlanta, USA. Detailed methodology for BINA is available in previous publications (17, 18). This paper presents quantitative and qualitative research on the experience of using 3D scanners for anthropometry. We conducted a time-motion study and qualitative interviews of anthropometrists, adopting a mixed-methods approach to provide a comprehensive assessment of experience. We collected data at the end of the BINA study (February 2017) to maximize anthropometrists' experience with AutoAnthro. We worked with anthropometrists to improve the relevance and application of our research (19), referring to Evidence Based Principles to Guide Collaborative Approaches to Evaluation (20) to inform our collaborative approach. BINA anthropometrists helped to design the research, developed tools, reviewed and revised manuscript drafts and approved the final manuscript.

6.2.1. Time-Motion

The study compared time required to take manual measurements versus 3D scanning using continuous observation based on milestone timing (21). We followed Suggested Time and Motion Procedures (STAMP) developed by Zheng et al (22) and developed and pretested the study protocol and tools. The lead researcher and one BINA anthropometrist defined all measurement tasks and developed cues for start/stop times based on those tasks. A single observer recorded time using a stopwatch in order to improve reliability. The intended sample size was a minimum of 22 children, based on achieving 0.9 power using a Paired T Test (α<.05) to detect a 30 second difference between measurement methods, which we considered a meaningful difference for efficiency of nutritional screening. The time motion study sample was a subsample of children who were measured in BINA, and we used purposeful selection to ensure that ~½ of the subsample was under 2 years of age.

We did not include the time to establish rapport or undressing because these activities were required for both scans and manual measurement. The protocol for manual measurement required undressing to undergarments while the scan based measurements required that the child be undressed to undergarments or to skin-tight leotard/shorts. We also did not measure the time required to set up equipment, but did include data entry in manual measurement time. Each child was scanned and measured twice by two different measurers, resulting in four sets of scans and four sets of manual measurements. A set of scans consisted of six one-second scans and a set of manual measurements included length or height, head circumference and arm circumference. To compare scans and manual measurements we timed the four sets as one unit, which simplified measurement by reducing the number of cues. Weight was not included in timing of manual measurements because we did not calculate weight from scans. The mean time for taking a single set of scan measurements was calculated by dividing the total time by four. To determine the time required for each manual measurement type (stature, HC, and MUAC) we observed a single measurer and took the mean from their two observations. We paused timing of measurements when measurers were interrupted. If the interruption was not related to the child's behavior we did not record the length of the interruption. In the case that a child became too upset to continue measuring and the measurer had to stop measuring to calm the child, we timed the interruption, and noted if it occurred during scans or manual measurements. Timing interruptions allowed us to calculate measurement time with and without interruptions; we felt that the two estimates were needed because pausing measurements may have been more common in BINA than in a household survey because of the facility setting and the absence of the primary caregiver. While timing measurements we also observed if a child cried and noted if crying occurred during scans, manual measurements, or both. We used SPSS 20 (IBM Corp., Armonk, NY, USA) for data analysis.

6.2.2. Interviews

The first author (JC) led the qualitative component of the study and had relevant training in the methods used in this study. He also supported training of anthropometrists and supervised data collection for the main BINA study. Before carrying out interviews the lead researcher had developed opinions on the new technology through experience with BINA; every effort was made to not influence the responses of interviewees. We used grounded theory from a constructivist point of view for qualitative design and analysis (23).

We sought to include everyone with extensive experience using the AutoAnthro System, which limited the intended sample to the five BINA anthropometrists, who all agreed to participate. In consultation with anthropometrists, we first conducted the written, in-depth interviews (IDI) and followed IDIs with a focus group discussion (FGD) that was facilitated by the lead researcher. One of the anthropometrists and the lead researcher developed a questionnaire with open-ended and probing questions on the identified categories of efficiency, invasiveness and the general user experience; with the latter category broadly covering the advantages and disadvantages of 3D imaging in comparison to manual measurements. The lead researcher and one anthropometrist independently coded IDIs line-by-line, and the lead researcher created code families and memos based on both sets of coded IDIs. We used ATLAS.ti 7 (Scientific Software Development GmbH, Berlin, Germany) for analysis.

Following the analysis of all individual, written interviews, we designed a semi-structured interview guide with open-ended questions for the focus group discussion. The primary purpose of the focus group discussion (FGD) was to clarify and expand on points raised in the written responses. The lead researcher facilitated the FGD which was held in a private conference room at Emory University. The proceedings were audio-recorded with a mobile phone and transcribed with Dragon NaturallySpeaking (Nuance, Burlington, MA, USA). The lead researcher coded the FGD line-by-line and revised code families generated from IDIs to incorporate the new information from the FGD. The lead researcher created a network map of code families to facilitate further memoing, and identified theories from the data. As authors, anthropometrists reviewed and revised findings, which functioned as a 'member check' to enhance trustworthiness of findings. We referred to the 2014 Standards for Reporting Qualitative Research from O'Brien (24) to report qualitative findings.

6.3 Results

6.3.1. Time-Motion

We observed and recorded measurement time for 27 children under five years of age. On average it took just over a minute (68 seconds) to complete a set of scans compared to over two minutes (135 seconds) for a set of manual measurements. The differences in measurement time between scans and manual measurement were statistically significant but did not differ by age group (Table 1). At the individual level, manual measurements took longer to complete for all children except one. For the child whose scans took longer the scanner was malfunctioning. There was little difference between stature, HC and MUAC for measurement time; with each measurement taking close to 40 seconds (Figure 1). Differences remained small after disaggregating by age group. For children under two years, the time for the various manual measurements ranged from 39 to 42 seconds, and for children over two years from 44 to 47 seconds.

Around 20% of children cried during manual measurements, and all of them were under two years of age (Table 2). Only one child cried during scans (4%), and that child also cried during manual measurements. Measuring was interrupted by the child's behavior on two occasions; both interruptions occurred during manual measurements and the average time of the interruption was 43 seconds. Including interruptions increased average manual measurement time by ~one second, increasing mean measurement time to 136 seconds.

6.3.2. Interviews

We completed five IDIs and one FGD, with participation in both from all five anthropometrists who had extensive experience using the AutoAnthro System. The five anthropometrists were all women with postsecondary degrees. After merging similar codes, we identified 96 action-oriented codes and 15 code families: time, cooperation, ease of use, staff, learning, invasiveness, caregiver, individual, child's age, clothing, experience, touch, safety, environment, and dependability. Table 3 presents the code families and selected, associated codes and quotations. We identified two core themes, or theories, from the data that related to favorable and unfavorable perceptions of using the AutoAnthro System: 'streamlined experience' and 'quality control'.

6.3.2.1. Streamlined Experience

Favorable perception of the 3D imaging system was dominant. The term 'streamlined experience' borrowed from one of the interviewees, who reported that *"scanning equipment…makes the process more streamlined."* We combined 'streamlined' with 'experience' to emphasize that streamlining could be applied to both physical equipment and the measuring experience, and also to highlight the importance of previous child experience.

"One of the major benefits…is that the device is extremely light, can fit in a small bag, and is very easy to operate. This is unlike the manual equipment, which is heavy and cumbersome…The scanning equipment also does not need to be sanitized… Further, the scanning equipment can double as…an entertainment device…"

The above quote highlights the physical characteristics of scanning equipment and how the use of a scanner effects other equipment needs. In addition to being smaller and lighter than a length board and a single piece of equipment compared to the multiple tools required for manual measurement, scanning reduced the need to carry additional supplies to sanitize the measuring equipment, and toys or other devices to encourage cooperation from children. Another anthropometrist stated that *"the 3D imaging device…eliminates the need for other resources,"* and referred to an additional advantage of the scanner doubling as an instrument to record measurements.

For an anthropometrist the experience of measuring begins with learning to use anthropometric tools, and all five anthropometrists commented that learning to take scans was easy, pointing out that *"besides having to adjust the box on the screen to fit the object/person…it is just like taking a picture*," and that there was "*not a lot of user input required to actually perform the scan."* For the most part the ease of learning to scan carried over to using the scanner in the field.

Most anthropometrists felt that in general taking scans was easier and faster than manual measurements. They pointed out that scanning saved time because they did not need to set up and sanitize equipment, or record measurements. They also felt that children were less fearful of scans than manual measurement because they related manual measurement to a painful visit to the doctor's office and scanning to "*getting their picture taken*." Anthropometrists felt that children were familiar with tablets and taking pictures, and that this familiarity made it easy to establish rapport. They also reported that confinement in the length board was a major source of distress for children. Overall AutoAnthro provided a 'streamlined experience'— it was easy to learn, the scanner itself was convenient, children did not experience stress, and taking scans was like taking a few pictures. However, all anthropometrists pointed out that taking scans was not always easy.

> *"I think that overall, the scanning technology is easier/faster/more convenient for children of all ages. If I were tasked with measuring children with either tool, I would want to have the scanning technology as my primary method, and have the traditional tools as a backup for cases where it wasn't feasible."*

The anthropometrist quoted above preferred scanning over manual measurements, but also felt scanning may not always be feasible. All anthropometrists reported that scanning was *"difficult, slower, and less dependable [than physical anthropometry] with an uncooperative/misbehaving child."*

6.3.2.2. Quality Control

Anthropometrists' view of scanning as a streamlined approach changed when capturing a good scan was out of their control. Scanning was dependent on child cooperation because movement affected the ability to capture high quality scans, and lack of cooperation was a common issue with children between the ages of 6 months and 3 years.

> *"I would say 3D scanning is more difficult, slower, and less dependable [than manual measurement] with an uncooperative/misbehaving child. Since we can't touch them while getting a scan, they can run or move around, making it impossible to capture a good scan."*

The problem of movement during scanning was exacerbated by two factors. First, the best solution to keep an active child still for the required second would be to physically hold them, but this was not possible during scanning because scan processing software required physical separation. Anthropometrists relied on various techniques to foster cooperation, but none of the techniques worked all of the time and sometimes anthropometrists *"gave up getting good scans and decided to settle for…subpar scans."* Second, anthropometrists were confident in identifying a good or bad scan, but for a scan that was somewhere in between good and bad, a "subpar scan," they were not certain if the scan was of adequate quality to process into accurate measurements.

Another less common situation when anthropometrists felt that capturing a quality scan was out of their control was when they experienced "software glitches." Anthropometrists knew that scanners could not function properly in direct sunlight because the scanner relies on infrared light that is washed out by direct sunlight, but they also reported reduced functionality under some "fluorescent lighting." Dim light did not cause any scanner problems and for the most part anthropometrists were able to move around the room and ensure "suitable lighting" through trial and error. This was not a problem when the anthropometrists could stay in one location within a facility for an entire day or multiple days, but became a challenge when they measured in a hospital setting where they had to move from room to room for each child; they considered finding "suitable lighting" a burden, in part because they could not always predict which lighting conditions constituted "suitable lighting." On occasion scanners did not function for an extended period of time, and while anthropometrists presumed light was the likely cause they could not identify the exact reason and sometimes referred to such situations as "software glitches."

> *"The scanning equipment is fairly reliable, but there have been times where there are issues with getting the camera to pick up the child or focus, which is very frustrating…there have been about 4 occasions where it took several minutes just to get one scan."*

At times anthropometrists were uncertain they would be able to complete a set of scans. With manual measurements there was little concern about completing measurements; uncooperative children could be held and there were no glitches with manual equipment. Additional qualitative findings on efficiency and invasiveness are integrated into the discussion.

6.4 Discussion

Four out of the five anthropometrists reported that scanning was faster than manual measurement and quantitative measurement showed that on average the time required for manual measurements was approximately two times longer than scans. During observation we found that nearly all crying episodes occurred exclusively during manual measurement, and in interviews four out of five anthropometrists indicated that children were more comfortable with scans. For the most part, the quantitative and qualitative components of this study were in agreement. Anthropometrist interviews provided further insights into efficiency, invasiveness and the user experience; including that increased efficiency and reduced invasiveness made scanning a 'streamlined approach' for most children, but that scanning was not easy for uncooperative children.

The vast majority of studies and research using 3D imaging for anthropometry did not include children under five years of age because imaging systems were not designed to handle movement. To our knowledge the AutoAnthro System is the first 3D imaging system designed specifically for full-body anthropometry of infants and young children. The only other 3D imaging system designed for young children is StarScanner (Orthomerica, Orlando, FL, USA), an approved medical device for measuring a newborn's head to design orthoses for cranial remolding (25). The AutoAnthro and StarScanner systems share the same capture strategy for handling movement — taking multiple scans of short duration and stitching them together. Our study showed that the capture strategy worked well for newborns, infants under six months of age, and children three years of age and over. However, anthropometrists felt that for up to ½ of children six months to three years of age it was difficult to get them to stay still long enough for multiple, one-second scans, and that often they settled for 'subpar scans.' Interestingly, in a previous publication we showed that in BINA the reliability of scan-derived measurements was not affected by the age of the child (17), which suggests that many of the 'subpar scans' were good enough and cooperation did not often have an effect on measurement quality. BINA anthropometrists took more than 6 scans and selected the 6 scans that they considered as the "best quality" for processing into each measurement. We do not know if cooperation affects scan-derived measurement quality when anthropometrists take only 6 scans. More research is needed to determine how often scans are of insufficient quality, and AutoAnthro should be improved to give anthropometrists confidence that they are capturing good quality scans. An ideal solution that is simple, but technologically complex would be to redesign the software to allow the anthropometrist or caregiver to hold the child during scanning. If quality is rarely affected by movement with the current software, as is suggested by reliability data, it may be sufficient to offer improved operator feedback that allows anthropometrists to distinguish between a "subpar scan" and a scan of adequate quality. Information on scan quality could be built into the software or provided through supervision. The need for additional feedback was expressed by an anthropometrist in the quote below.

> *"…we have not had sufficient feedback to know if all our submitted images are adequate. We do get this type of feedback on our manual measurements. We can see if our first and second set of measurements are close and we can compare…to that of our partner…The times that we tested this and found that we were measuring consistently too big or too small, we could correct our technique.*

Before BINA we did not have data on the reliability and accuracy of scan-derived anthropometry; now that we have such data it is possible to develop metrics of scan quality in relation to anthropometry.

The current cost of the 3D scanner used in the AutoAnthro System (USD $379 (26)) is more than a length/height board (USD $122 (27)). The requirement to attach the 3D scanner to a cell phone or tablet adds cost to the imaging system, but electronic data capture is becoming more common in clinics and surveys, and the added cost of a mobile device is likely offset by eliminating the need for paper and data entry. In this study we found that 3D imaging brought efficiency gains related to training, staff, and measurement time that may help to further offset increased costs. We spent one day on 3D imaging training, and anthropometrists felt that learning to use the scanner was easy. Our previous findings on reliability suggest that the one day training on 3D imaging was sufficient because scan-derived measurements were reliable and between-measurer reliability was the same as within-measurer reliability (17). Reduced training time could offer savings over traditional anthropometry, which relies on one to two week trainings of anthropometrists to achieve quality results. In addition to substantial training, manual anthropometry requires the use of a trained assistant. Anthropometrists felt that an assistant was needed for 3D imaging to help position the child, but that *"it doesn't necessarily have to be someone who is trained."* In some settings 3D imaging may be able to rely exclusively on the caregiver to act as an assistant, reducing staff needs. The garment industry saw 3D imaging as a way to reduce the time required to carry out large sample size surveys with up to 40 separate manual measurements of each individual (1). We found that 3D imaging took less time than the three manual measurements included in our study. For a household survey, the small difference in measurement time between scans and manual measures may not represent a meaningful difference for efficiency because much more time is spent traveling from house to house. For large-scale screening however, saving a minute per child could be considered important. However, in the health sector only child weight and length/height are commonly measured, with head circumference limited to newborns and infants, and MUAC used as an alternative to weight for length/height. Additional manual measurements are generally not used because of the difficulty and burden of measuring. Compared to the one or two common, manual measurements 3D imaging does little to reduce measurement time, but it does provide an opportunity to develop novel anthropometric indicators that are not feasible with manual measurement and that may be better predictors of outcomes of interest. Efforts to create new indicators based on 3D measures has already started, with the development of the Body Volume Index and the Health Index (12, 14). It is difficult to quantify the future value of new anthropometric indicators. Portability and reduced invasiveness are additional but important advantages of 3D imaging that are difficult to assign value to. The smaller dimensions and reduced weight (<0.5 kg) of the AutoAnthro System lessen the burden on anthropometrists and may reduce transportation costs when compared to a typical, wooden length/height board (7.7kg (27)), and anthropometrists reported that children experienced less stress during 3D imaging. After additional research is carried out and the scanning protocol is finalized, our findings can help to design a comprehensive costing study.

An important strength of this study is the research design; we used a mixed-methods, collaborative design that increased the relevance and trustworthiness of findings. However, there are some limitations that need to be considered when interpreting or generalizing the findings. We did not consider the time needed to process scans into measurements because the imaging system was designed to be fully automated. However, anthropometrists had to select the 12 best scans when they took extra scans, which was a frequent occurrence, and selection took a substantial amount of time. We expect that an updated version of AutoAnthro will be fully automated, but with the current version we underestimated measurement time by not taking into account selection and deletion of scans. Both qualitative and quantitative findings were based on a small sample size. According to the anthropometrists, scanning took longer for uncooperative children compared to cooperative children. We therefore expected to find a difference in scanning time by age because most uncooperative children were under two years of age, but this was not the case; the lack of differences may however be due to the small sample size for the time motion study. For the interviews there were only five anthropometrists with sufficient experience using the AutoAnthro System, and all of them had post-secondary education and were familiar with electronic devices such as tablets and smartphones. Future research on user experience could include multiple focus group discussions to get input from a larger number of anthropometrists, and findings from this study should not be extrapolated to anthropometrists or children/caregivers with less education and/or limited experience using similar technology. The sensitivity of scanners to light may be more problematic when scanning outside of a building or with frequent movement from house to house, and if caregivers and children lack previous experience with mobile devices they may react differently to the technology. Anthropometrists reported that some primary caregivers did not consent to their child participating in the study because of privacy concerns and that the safety of the 3D scanner was a common concern of caregivers during recruitment. In BINA we did not use a formal sampling frame and it was not possible to determine the percentage of caregivers that refused consent. An additional limitation is that we did not collect qualitative data directly from parents/caretakers or children. Since we conducted most of the measurements in day care and hospital settings, caretakers were not often present during data collection and the majority of children in the study were not old enough to be interviewed.

In a previous publication, we described the need for further research on AutoAnthro to replicate reliability findings and to remove systematic inaccuracy (17). This study further supports the need for additional research before we can make a recommendation for the widespread use of AutoAnthro in surveys or regular nutritional assessment. Specifically, research to develop scan quality control mechanisms is needed. The scanner needs to be tested in a household or community setting with different lighting conditions; and in a population that is not familiar with similar technology. In addition to providing information on different lighting conditions, a household study could determine the likelihood of a caregiver refusing to have their child scanned. As this study showed, it is important that future studies include a qualitative component to provide a comprehensive evaluation. Our findings on efficiency, invasiveness and the user experience could vary dramatically in a different setting. For additional qualitative research it would be helpful to carry out a study with caretakers present; the caretakers themselves may provide valuable insights into invasiveness and the general experience with the technology; and it may also be possible to get feedback from older children.

6.5 Conclusions

In this study we found that anthropometrists were not yet ready to completely abandon traditional, manual equipment for 3D scanners. For most children under five years of age 3D imaging was an efficient and non-invasive way to capture anthropometric data. Revising the AutoAnthro System to address anthropometrists' concerns on capturing good quality scans of uncooperative children should help to facilitate widespread use of 3D imaging for child anthropometry in the health sector.

6.6 Figures and Tables

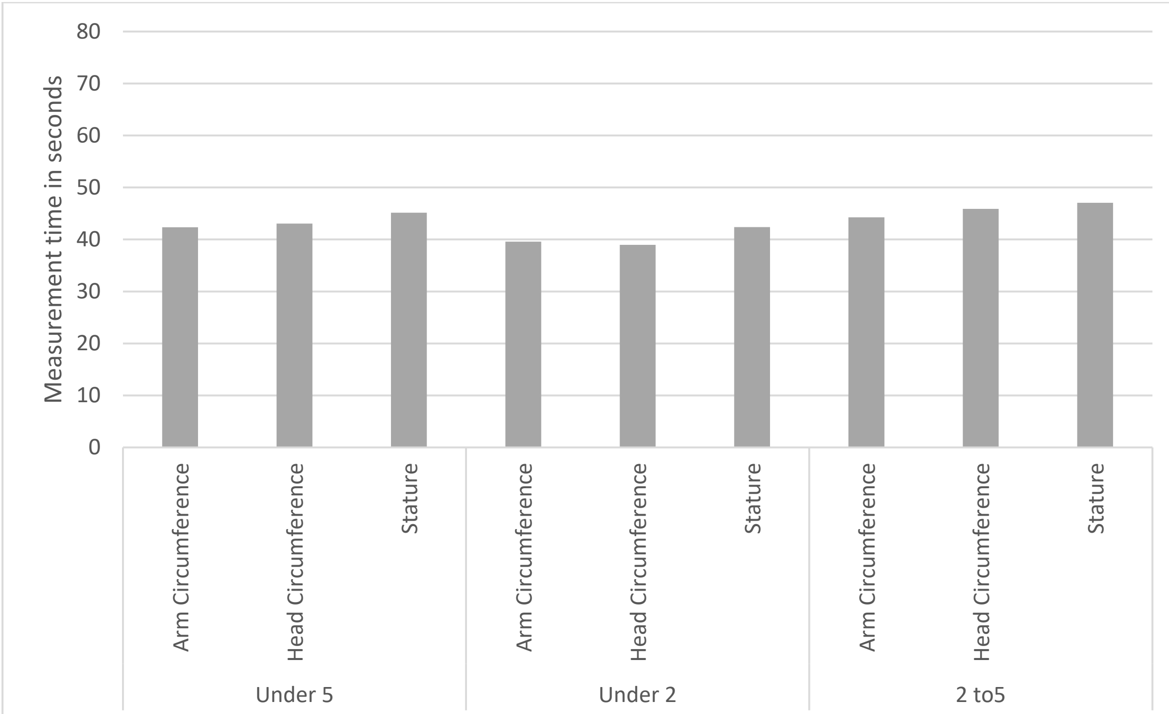


Figure 6-1 Mean measurement time and standard deviation for manual measurements completed by the first measurer, BINA 2017

Table 6-1 Difference in the time required to complete scans and manual measurements, BINA 2017

Age	n	Average time to complete scanning (seconds)		Average time to complete manual measurements (seconds)		Paired Samples T-Test (difference and 95% confidence intervals (CIs) in seconds)			
		Mean	SD	Mean	SD	Difference (scan-manual)	95% CI Lower Limit	95% CI Upper Limit	Significance
Under Five	27	68	29	135	22	-67	-80	-54	<.001
Under Two	11	63	23	121	20	-58	-80	-35	<.001
Two to Five	16	71	32	144	18	-73	-91	-56	<.001

Table 6-2 Crying episodes and interruptions caused by non-compliance during scans and manual measurement, BINA 2017

Age	n	Number of children crying during scans or manual measurement	Number of children crying during scans	Number of children crying during manual measurement	Number of interruptions[a]	Average time of interruption (seconds)
All	27	6	1	6	2	43
Under 2	11	6	1	6	2	43
2 to 5	16	0	0	0	0	n/a[b]

[a] Both interruptions occurred during manual measurement

[b] Not applicable because no interruptions in this age group

1 Table 6-3 Results of coding and memoing of anthropometrists' interviews, BINA 2017

Code Family	Selected Codes (underlined) & Selected Quotations	Summary of Memos (code families in bold)
Time	Cooperation Same for Both: "Generally, if a mild mannered child is cooperative for one set of measures they will be for the other, but the 3D scanning takes LESS TIME than the physical." Child Moving: "There are "sweet spots" in terms of age that makes the scanning faster than traditional measurements. Newborns and young infants… can be scanned faster…"	A major driver of the time required to complete measurements was child **cooperation**, which itself was driven by the **age** and temperament of the child, and **invasiveness** of the measurement method. Most anthropometrists reported that scanning took less time than manual measurements when children were cooperative. All anthropometrists reported that scanning took more time for uncooperative children compared to cooperative children, and some felt that scans took more time than manual measurements for children six months to three years of age.
Cooperation	Crying: "…reactions during physical measurements occurred while children under two were getting their length measurement taken. Most children, even the most compliant, did not enjoy the length board and usually cried, screamed, or tried to stand up." Assessing Blame for No Cooperation: "…we just didn't start anything on them because they were so uncooperative, so I wouldn't blame that on the scans or the physical measurements. I would kind of blame it on the whole process. Distracting Child: "Except for height/length, there is not an exact [full body] pose…for physical measurements. The measurer can move around the child to do the measurements. It is easier to distract a child during the physical measurement process. They can watch a video or play a game on the iPad…"	The method of measuring, scanning versus manual, was not the main determinant of cooperation. However, length was consistently reported to be particularly difficult. Anthropometrists viewed the child's temperament as important, and viewed **child's age** as the best predictor of cooperation. Anthropometrists reported that removal of **clothing** and "stranger anxiety" could initially cause distress of the child, leading to poor cooperation and refusal before attempting measurement in some cases. Some anthropometrists reported that once measurement began refusals occurred exclusively during scanning, while others reported refusals only occurred during manual measures. During measurement distraction was the main strategy used by anthropometrists to foster cooperation, and there was consensus that showing videos was the best distraction tool. Scanning uncooperative children was especially challenging because the child was unable to move, anthropometrists could not **touch** the child, and distraction was more difficult.
Ease of Use	Easy to Carry: "One of the major benefits…is that the device is extremely light, can fit in a small bag, and is very easy to operate. This is unlike the manual equipment, which is heavy and cumbersome during transportation." Using both tools: "I think that overall, the scanning technology is easier/faster/more convenient for children of all ages. If I were tasked with measuring children with either tool, I would want to have the scanning technology as my primary method, and have the traditional tools as a backup for cases where it wasn't feasible." Child Moving: "I would say 3D scanning is more difficult, slower, and less dependable with an uncooperative/misbehaving child. Since we can't touch them while getting a scan, they can run or move around, making it impossible to capture a good scan."	Anthropometrists commented that the physical characteristics of the scanner, small and lightweight, made it easy to use. **Learning** to use the scanner was easy according to anthropometrists, who also commented that the scanning equipment was sturdy and did not require sanitization. Anthropometrists did not consider charging the scanner battery to be a big burden, but did report that forgetting to charge led to data collection delays on occasion. There was no reported potential for harm to the child from the scanners, but anthropometrists reported needing to explain the **safety** of scanning technology to caregivers. Children's previous **experience** with cameras facilitated easy use of scanners, but the requirement to remove **clothing** was a burden. Most anthropometrists reported that completing all required scans was generally easy, but that some 'trial and error' was necessary and that scanning became difficult in specific circumstances. **Environment**, specifically lighting (attributed to both natural and fluorescent light), affected scanner functionality and made data collection more difficult. The biggest reported challenge to

		use the scanners was getting an uncooperative child to stay in position long enough to obtain adequate scans without being able to **touch** the child.
Staff	Not Needing Trained Staff: "I think it's helpful to have two staff for manual measurements. It's helpful for scans to have a staffer and someone else to position the kid. It doesn't necessarily have to be someone who is trained." Seeing Different Parts: "I think both always need two if you want to be accurate. For height and length someone has to be watching one end of the body. If you're doing scans, unless you have a perfect kid who was understanding your verbal directions…, [a single operator] would have to walk over change their arms, come back scan."	Reported staff needs varied from one to three depending on the measuring method and **age** of child. For uncooperative children under two years of age anthropometrists reported needing three people to get an accurate length measurement. For manual measures at least two trained staff were necessary because height measurement requires simultaneous viewing of different parts of the body to ensure correct positioning. For scanning there was some agreement that an assistant was needed for most children and helpful for all children, but the assistant did not necessarily need to be formally trained. Some anthropometrists reported that with a cooperative child who followed instructions some manual measurements and scanning could be completed by a single measurer. For both manual measures and scanning anthropometrists viewed the use of an assistant as important for reducing measurement **time**.
Learning	Working by Trial and Error: "Learning to use 3D imaging was relatively easy. Besides having to adjust the box on the screen to fit the object/person…it is just like taking a picture. Most of the lessons were learned through trial and error…" Being Confident in Measurement: "Instructions on using the 3D imaging was pretty straightforward…I'm not sure if all of the measurers' questions concerning what constitutes a 'good scan' were ever completely answered. Though it was pretty obvious on what signified a 'bad scan."	There was unanimous agreement that 3D imaging was easy to learn; it was like taking a picture. The custom software for scanning did not require much user input. However, trial and error was necessary during data collection to learn how to deal with various circumstances. For example, anthropometrists found that it was not possible to scan in hallways or to use two scanners at the same time on a single child. For the most part anthropometrists learned how to ensure that **environment** did not affect scans. However, at the end of data collection anthropometrists still did not feel that they could recognize a "good scan," and they could not perfectly predict when lighting would affect scans.
Invasiveness	Receiving Medical Care: Children generally detest getting their recumbent length taken. If they are old enough, they might think they are getting a shot when we do MUAC, even when we explain what we're doing. They often think whatever physical measurements we are going to do will hurt – because they associate us with medical professionals. They tend to not have these fears with scans. Being Scared: For traditional measuring a lot of children first tend to be a little afraid…because our process is very similar to what they experience when they visit the doctor's office and they associate going to the doctor with getting painful shots. After seeing that what we are doing with them is not painful, I've noticed that most children are pretty relaxed and happy to be measured. For 3D imaging, children seem to have more fun and are excited to do the poses (sometimes too excited). Because it's kind of like taking a photograph and there is way less touching involved, I think most children are way more comfortable with this method.	Anthropometrists defined measuring invasiveness as causing the child to be "uncomfortable," "anxious," or "distressed;" and reported related behaviors of "crying," "screaming," or "moving away." Removal of **clothing** was seen as invasive for some older children, and this was related to "stranger anxiety." Nearly all anthropometrists reported that children were more comfortable with scanning because they were used to having a picture taken. Previous **experience** also affected manual measurement; all anthropometrists reported that children related manual measurements to a doctor's visit, with MUAC being related to getting shots. All anthropometrists reported that length caused the most distress, and the sense of confinement was cited by some as the underlying reason. **Touch**ing occurs for manual measurements and in some cases children were anxious about being touched, while for others touching was a source of comfort. Anthropometrists also considered the caregiver's reaction when considering invasiveness. Anthropometrists reported that caregivers may be more comfortable with manual measures because they are already familiar with them, and because there is an aversion to taking what appears to be a picture when the child is undressed. However, negative reaction from **caregivers** was reported for the length measurement.

Caregiver	<u>Parents Make Harder</u>: "I noticed that children tend to respond negatively to any of the measures when there was a parent around." <u>Discomfort with Stranger</u>: Many times, the child/infant did not like being touched by strangers (us). The whole process generally went better when we had a caregiver assist. <u>Feeling Awkward</u>: "…showing a scan to the caregiver was a way to reassure them that the 3D image was much different than a photograph and that the child's identity and privacy was protected."	Multiple anthropometrists felt that the presence of a caregiver made measurement more difficult, but all agreed that undressing the child was easier with a caregiver present. Some anthropometrists reported that uncooperative behavior of the child during measurement was more common when a parent was present. Some anthropometrists felt compelled to show caregivers the scan of the child to reassure them that it was not an identifiable photograph. Anthropometrists reported that caregivers expressed that previous manual measures of their child in the doctor's office were inaccurate, and that they were hopeful the scanner could provide accurate measurement.
Individual	<u>Dealing with Language Barrier:</u> "I think it is pretty easy for children to understand what we need from them in order for us to get good scans. Challenges only occur when the child is really active…, or when there is some sort of barrier in communication." <u>Assessing Child Temperament</u>: "Being able to detect the child's temperament and learning style early on did contribute to the time it took to secure measurements. Each child responded to different distraction techniques and games in various ways. Identifying the type of child being measured early on was often helpful in decreasing measuring time."	Anthropometrists highlighted individual child characteristics when discussing measurement efficiency and ease of use, referring primarily to child "temperament." Specific behaviors that made measuring more difficult and time consuming were: being active or unable to stay still, and seeking attention. Distraction techniques had to be adapted to the individual child. Over-activity and attention seeking were viewed as more problematic for scanning because of the inability to **touch**, and some anthropometrists believed that scanning exacerbated attention seeking. Multiple anthropometrists discussed language as a barrier for efficient scanning because of the inability to communicate positioning to the child or caregiver who acted as an assistant when English was not their first language. Language barriers affected scans and manual measurements because a lack of understanding made the child more afraid.
Child's Age	<u>Getting Usable Scans</u>: "I think the age of the child does have an effect on which measure is faster. Most of the 3 and 4 year olds were easy to scan and measure. Children that were old enough to crawl (about 9 months) and under 3 took longer to scan because we had a hard time keeping them still enough to capture usable 3D images." <u>Child Lacking Awareness</u>: "Clearly young infants…are not aware of what is going on…they don't seem to have a reaction to either the scanning or the physical measurements. Children over about 6 months become more difficult to manage. They may not want to stay in the position…for the scanning, and may resist being touched for the physical measurements. Children well over 3 years old frequently do understand…and can be quite cooperative."	All anthropometrists agreed that the age of the child was the largest determinant of the speed and ease of measuring. Infants under 6 months and children older than three years were the easiest to measure. When infants start to turn over and crawl the movement makes measuring more difficult. At one year of age awareness increases and children can become "knowingly uncooperative." Child strength increases with age and children become harder to physically manipulate, which can make measuring more difficult from one year of age until the age at which children are better at following directions, 2.5 to three years of age. Within the more difficult age group of 6 months to 3 years, children 12-24 months were particularly challenging because they did not like to lie down and are strong enough to resist. While both manual measurement and scanning were more difficult for the middle age group, the inability to touch the child made scanning more difficult for this age group.
Clothing	<u>Undressing a Child</u>: "Older children (36+months) are the age range that generally have the most concern about being undressed. The process of getting the child undressed and into another form of covering has taken up to 10 minutes, multiple visits, assistance from adults the child is comfortable with, and sometimes has required the case to be lost. Regardless of age of the child, parental figures have lost their willingness to consent due to the requirement to undress for scanning."	All anthropometrists reported that undressing the child was a challenge. Undressing caused distress before measuring began. Anthropometrists related discomfort with undressing to "stranger anxiety." Older children were more reluctant to undress. One anthropometrist felt that undressing caused children to relate measuring to experience at the doctor's office. Some anthropometrists reported that they themselves felt awkward undressing children, but that it became easier as the study progressed. Anthropometrists, who also recruited for the study, reported that some caregivers were hesitant or refused to consent to the study because children would be undressed.

Experience	<u>Receiving Medical Care</u>: "The MUAC measuring tape seemed to remind children of the tourniquet applied to the arm before shots are administered." <u>Receiving Medical Care</u>: "A pen or marker is typically used to mark the midpoint which can sometimes be confused as a needle for taking a shot and can have adverse effects on the child's behavior." <u>Comfortable because Familiar</u>: "I believe that children are more comfortable with a tape measure and the measuring board. These are items they have seen before and have some understanding of how they work." <u>Taking a Picture</u>: I think the older children actually enjoy doing the scans, because they think they are being photographed doing poses.	All anthropometrists commented that the previous experience of the child affected the measurement experience. One anthropometrist commented that children were taught not to undress for strangers, and another reported that undressing reminded children of visiting the doctor. One anthropometrist felt that children were more comfortable with manual measurement because they were familiar with the equipment. The most commonly reported beliefs from the anthropometrists were that children related scanning to having their picture taken and manual measurement to going to the doctor's office. All anthropometrists said that manual measurement equipment made children relate the experience to going to the doctor, sometimes causing distress and uncooperative behavior. The children themselves made comments that convinced anthropometrists that they thought it was a doctor visit. Children were familiar with tablets and phones; and all anthropometrists agreed that older children related scanning to taking a picture. For the most part the idea of taking a picture made children more cooperative, but some felt it exacerbated attention seeking in some cases. The scanner made a "clicking" sound, which may have reinforced the idea of taking a picture.
Touch	<u>Holding the Child</u>: "Being able to hold the child in order to complete measurements (on the length board for example), certainly makes the process much faster. Not being able to hold the child frequently makes the scan process take much longer…" <u>Holding the Child</u>: "While capturing scans it would be extremely helpful if the child could be touched. This would aid in keeping them still and in their proper poses." <u>Getting Frustrated</u>: "…the experience of taking scans with the really uncooperative child is so emotionally infuriating…It's pretty common for it to be difficult to do physical measurements where the kids scream and you can just tune them out, but when you try to take scans and they're running around it's so frustrating and it makes you really upset...[It happened] Probably once a day."	Anthropometrists reported that some children were sensitive to being touched by strangers. For children that were sensitive to touch manual measurement was more distressing to the child than scanning, but anthropometrists did not consider touch sensitivity a big issue. The larger issue with touch was the inability to touch children during scanning, which made positioning the child and keeping the child still much more difficult and time consuming. Through trial and error anthropometrists started to use long spoons – during scanning the child could hold one end while the anthropometrist held the other end of the spoon, and it did not affect the quality of the scan or the ability to process the scan. The use of spoons helped mitigate the impact of not being able to touch the child, but it did not always work; and it was common for scanning to take longer for active children that did not follow instructions. For some anthropometrists the inability to physically restrain children during scanning was a frequent source of frustration. Others reported that feelings of frustration were not so frequent.
Safety	<u>Children Playing with Equipment</u>: "The board does not move smoothly, so sometimes it can bump a child on the head as it snaps into place. The other main issue with the height board is that children often like to try to measure their own head, grabbing the moving part of the board and pulling it down. If the measurer isn't quick enough, this results in them hitting themselves pretty hard on the head. I don't think I've ever seen a child become upset by this though."	Anthropometrists did not report any harm to a child from scanning or manual measuring. The only reported safety concerns of the anthropometrists were that moving pieces of manual equipment could potentially hurt children, and anthropometrists did sometimes worry about hurting the child when physically manipulating them into position for manual measurement. For scanning, sanitization of equipment was not necessary, and one anthropometrist mentioned that there was less chance for spreading pathogens during scanning because there was less physical contact. Anthropometrists reported that some caregivers showed concern over 3D scanning "being harmful to the child internally," and over taking pictures of children without clothing.

Environment	Other Children Influence Cooperation: “A factor to consider for both measuring types is the environment around the children. If it is cold in the room, both processes can be very uncomfortable. Some children do better with other peers around and some do not. Even things like having other toys in the room can be a distraction for both processes and increase the time spent on each case.” Lacking Certainty: “I still haven't figured out with the lighting how changing it affects the scans because it's not consistent.” Selecting Scan Location: “[Scanner malfunction from lighting] probably happened at every site when we first got there. At Midtown we were in different rooms every time and we had to figure that out every single time.”	Anthropometrists mentioned some environmental concerns that affected both manual measurements and scanning, such as cold causing children to be uncomfortable and objects or other children in the room affecting cooperation. There were additional environmental factors that were reported only in relation to scanning. A flat surface was necessary for scanning, as was adequate space. Anthropometrists found that they needed enough distance between themselves and the child to capture the entire child in a scan, and that narrow spaces (such as a hallway) would make the scanner malfunction. Lighting was the most commonly mentioned environmental factor, and it seemed to be the hardest factor to account for. Anthropometrists reported that both natural and fluorescent light affected scans. At the end of data collection anthropometrists still did not always understand why light was causing scanner malfunction and could not always predict where lighting was appropriate for scans. For the most part lighting was not viewed as a big problem; anthropometrists would identify an appropriate place to scan children at each location and stay in that location. At one site anthropometrists had to move from room to room and this was the site where lighting presented the biggest problem for scanning.
Dependability	Lighting Affects Scan: “Overall, I think that the equipment for the physical measurements were more reliable and consistent because we didn’t have to worry about external factors such as lighting or space interfering with these measurements.” Lacking Certainty: “There were times that my scans had interference that I couldn’t determine the source. Was it poor lighting? Reflection from a surrounding material?” Experiencing Glitches: “On many occasions the measuring tapes have become damaged, but they are cheap… The scanning equipment is fairly reliable, but there have been times where there are issues with getting the camera to pick up the child or focus …there have been about four occasions where it took several minutes just to get one scan on a child...”	Anthropometrists rated manual equipment as the most dependable because it was sturdy and consistent. With manual equipment there was no concern of external, environmental factors affecting measurement. Anthropometrists reported that measuring tapes frequently broke, but this was easily dealt with by using replacements. Anthropometrists viewed scanners as generally dependable, but there were exceptions. Scanners were viewed as “surprisingly sturdy.” There were no reported instances of 3D scanners getting damaged or breaking. Anthropometrists reported that charging the scanner and iPad was not a big burden and only took an hour, but sometimes operators forgot to charge in the evening and this caused delays in data collection. The main reason scanners were rated less dependable than manual measurement is that they did not always function properly. Anthropometrists reported experiencing “glitches” that caused delays in data collection. Malfunctioning was frequently attributed to lighting and in every location anthropometrists spent time to find a spot with appropriate lighting. In some cases anthropometrists could not determine the cause of scanner malfunction. Anthropometrists also highlighted that the dependability of the scanner was dependent on the child staying still.

2

References (Chapter 6)

1. Jones PR, West GM, Harris DH, Read JB. The Loughborough anthropometric shadow scanner (LASS). Endeavour. 1989;13(4):162-8.

2. Robinette KM, Daanen H, Paquet E, editors. The CAESAR project: a 3-D surface anthropometry survey. Second International Conference on 3-D Digital Imaging and Modeling (Cat NoPR00062); 1999 1999.

3. Wells JC, Treleaven P, Cole TJ. BMI compared with 3-dimensional body shape: the UK National Sizing Survey. Am J Clin Nutr. 2007;85(2):419-25.

4. Kim JY, You JW, Kim MS. South Korean anthropometric data and survey methodology: 'Size Korea' project. Ergonomics. 2017:1-11.

5. Charoensiriwath S, editor A real-time data monitoring and management system for Thailand 2019 first national sizing survey. PICMET '08 - 2008 Portland International Conference on Management of Engineering & Technology; 2008 27-31 July 2008.

6. Kuehnapfel A, Ahnert P, Loeffler M, Broda A, Scholz M. Reliability of 3D laser-based anthropometry and comparison with classical anthropometry. Scientific reports. 2016;6:26672.

7. Pepper MR, Freeland-Graves JH, Yu W, Stanforth PR, Cahill JM, Mahometa M, et al. Validation of a 3-dimensional laser body scanner for assessment of waist and hip circumference. Journal of the American College of Nutrition. 2010;29(3):179-88.

8. 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;10(3):e0119430.

9. Wang J, Gallagher D, Thornton JC, Yu W, Horlick M, Pi-Sunyer FX. Validation of a 3-dimensional photonic scanner for the measurement of body volumes, dimensions, and percentage body fat. Am J Clin Nutr. 2006;83(4):809-16.

10. Garlie TN, Obusek JP, Corner BD, Zambraski EJ. Comparison of body fat estimates using 3D digital laser scans, direct manual anthropometry, and DXA in men. Am J Hum Biol. 2010;22(5):695-701.

11. Adler C, Steinbrecher A, Jaeschke L, Mahler A, Boschmann M, Jeran S, et al. Validity and reliability of total body volume and relative body fat mass from a 3-dimensional photonic body surface scanner. PloS one. 2017;12(7):e0180201.

12. Barnes R, editor The Body Volume Index (BVI): Using 3D Scanners to Measure and Predict Obesity. International Conference on 3D Body Scanning Technologies; 2010; Lugano, Switzerland: Hometrica Consulting.

13. Loffler-Wirth H, Willscher E, Ahnert P, Wirkner K, Engel C, Loeffler M, et al. Novel Anthropometry Based on 3D-Bodyscans Applied to a Large Population Based Cohort. PloS one. 2016;11(7):e0159887.

14. Lin JD, Chiou WK, Weng HF, Fang JT, Liu TH. Application of three-dimensional body scanner: observation of prevalence of metabolic syndrome. Clin Nutr. 2004;23(6):1313-23.

15. Kuehnapfel A, Ahnert P, Loeffler M, Scholz M. Body surface assessment with 3D laser-based anthropometry: reliability, validation, and improvement of empirical surface formulae. Eur J Appl Physiol. 2017;117(2):371-80.

16. Weathers WM, Khechoyan D, Wolfswinkel EM, Mohan K, Nagy A, Bollo RJ, et al. A novel quantitative method for evaluating surgical outcomes in craniosynostosis: pilot analysis for metopic synostosis. Craniomaxillofac Trauma Reconstr. 2014;7(1):1-8.

17. Conkle J, Ramakrishnan, U., Suchdev, P., Flores-Ayala, R., Alexander, E.,Martorell, R. Accuracy and Reliability of a Low-Cost, Handheld 3D Imaging System for Child Anthropometry. Manuscript in Preparation. 2017.

18. Conkle J, Ramakrishnan, U., Suchdev, P., Flores-Ayala, R., Martorell, R. Methods and quality of manual anthropometry in the 3D Body Imaging for Nutritional Assessment Study (BINA). Manuscript in preparation. 2017.

19. Israel BA, Parker EA, Rowe Z, Salvatore A, Minkler M, Lopez J, et al. Community-based participatory research: lessons learned from the Centers for Children's Environmental Health and Disease Prevention Research. Environ Health Perspect. 2005;113(10):1463-71.

20. Shulha LM, Whitmore, E., Cousins, J.B., Gilbert, N., Al Hudib, H. Evidence Based Principles to Guide Collaborative Approaches to Evaluation: Technical Report. Ottawa: Centre for Research on Educational and Community Services, University of Ottawa; 2015.

21. Lopetegui M, Yen PY, Lai A, Jeffries J, Embi P, Payne P. Time motion studies in healthcare: what are we talking about? J Biomed Inform. 2014;49:292-9.

22. Zheng K, Guo MH, Hanauer DA. Using the time and motion method to study clinical work processes and workflow: methodological inconsistencies and a call for standardized research. J Am Med Inform Assoc. 2011;18(5):704-10.

23. Charmaz K. Constructing grounded theory. London ; Thousand Oaks, Calif.: Sage Publications; 2006. xiii, 208 p. p.

24. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med. 2014;89(9):1245-51.

25. Ifflaender S, Rudiger M, Koch A, Burkhardt W. Three-dimensional digital capture of head size in neonates - a method evaluation. PloS one. 2013;8(4):e61274.

26. Occipital Inc. Structure Sensor Sensors [Available from: https://store.structure.io/store?_ga=2.30735295.1074086949.1503324024-1021633441.1503324024.

27. United Nations Children's Fund. Supply Catalogue Product Details [Available from: https://supply.unicef.org/unicef_b2c/app/displayApp/(cpgsize=5&layout=7.0-12_1_66_69_115_2&uiarea=2&carea=50D129C26B9F08F2E10000009E710FC1&cpgnum=1&cquery=length+board&citem=50D129C26B9F08F2E10000009E710FC14EBA0434BBC20F68E10000009E71143E)/.do?rf=y.

Chapter 7 . Discussion

7.1. Summary of Findings

In Chapter 3 we found that the quality of manual measurements in BINA was excellent, and similar to the quality of anthropometric data used to develop the WHO Child Growth Standards. We attributed high quality to vigorous training, motivated and competent field staff, reduction of non-measurement error through the use of technology, and reduction of measurement error through adequate monitoring and supervision. We concluded that BINA offered a valuable evaluation of 3D imaging for child anthropometry because there was comparison to gold-standard, manual measurements. In Chapters 4 and 5 we found that measurement reliability of repeated scans was excellent, and similar to manual measurement reliability for stature, HC and MUAC. However, we found systematic bias when analyzing accuracy — 3D imaging overestimated stature and HC and underestimated MUAC. We hypothesized that human/equipment error in manual measurement of MUAC, and scan processing software that did not correctly account for soft tissue compression and body positioning during manual measurement of HC and stature respectively could have caused systematic inaccuracy. After adjusting measurements to remove systematic bias, 3D imaging yielded mean z-scores, z-score standard deviations (SD), and prevalence below or above z-score SD cutoffs that were similar to manual measurements. Based on a cutoff of one SD, specificity of adjusted, scan-derived measurements was excellent for all measures, and sensitivity was good to excellent for all measures. In Chapter 6 we showed that for cooperative children anthropometrists considered the use of AutoAnthro an easy, 'streamlined experience,' but with uncooperative children or when experiencing 'software glitches' anthropometrists felt that capturing a good quality scan was difficult and out of their control. From the time motion study we found that scanning took less time than a set of manual measurements (stature, HC, and MUAC) and we observed that crying was more common during manual measurement. Anthropometrist interviews confirmed that 3D imaging was less stressful for children than manual measurement.

7.2. AutoAnthro anthropometry quality in relation to other 3D imaging systems

7.2.1. Reliability

Similar to our findings, previous research of range imaging systems concluded that scan-derived anthropometry was reliable (1-5). However, it is difficult to directly compare our reliability findings with previous research because studies used different analytical methods. Most studies only reported correlation coefficients and each study used a different type of coefficient. Correlation coefficients are difficult to interpret because they are highly dependent on the amount of variation in the sample. If you measure a wide age group with high variation of size, a high correlation coefficient can say more about sample variation than measurement reliability. In the nutrition research community it is common to use the technical error of measurement[1]. The International Society for the Advancement of Kinanthropometry (ISAK), an organization that certifies anthropometrists, requires intra-observer %TEM of <2% for a level one anthropometrist and <1.5% for level two certification during examination (5). One of the studies that used Kinect scanners measured %TEM and found that when measuring stationary cylinders their 3D imaging system satisfied reliability requirements for level one anthropometrist accreditation for small cylinders and level two for large cylinders (5). In our research AutoAnthro intra-observer %TEM from single scan sessions was good enough to achieve level two anthropometrist certification for stature and head circumference in all age groups, but MUAC reliability was not high enough for accreditation. With repeated-scan AutoAnthro intra-observer %TEM would be lower, and would likely qualify for level one anthropometrist accreditation for MUAC. In Chapters four and five we found that when using repeated measures AutoAnthro measurement reliability was similar to gold standard manual measurement, and sufficient for prevalence estimation, growth monitoring, and individual classification.

[1] Technical error of measurement is not well known outside of nutrition research, and it is not included in common software for statistical analysis, such as SPSS and Stata. Technical error of measurement is very similar to one standard deviation of the mean of the absolute difference between two measurements. If the nutrition community used the mean of the absolute difference with standard deviation and a 95% confidence interval, reliability results would be easier to calculate and more meaningful than TEM for people outside of the nutrition community.

7.2.2. Accuracy and Classification

The ability of a 3D imaging system to produce correct measurements is dependent on the hardware (scanner) and software (computer program to process scans into measurements). When interpreting findings on accuracy both system components should be considered. The scanners and processing software used in national sizing surveys produced measurements that were accurate enough for industrial uses. When the same or similar scanners were tested in health sector research that used a variety of processing software some studies found accurate measurement (6), but many found systematic inaccuracy when looking at average bias —3D imaging produced higher average estimates than manual measurement for height (+~0.8 cm, (2)) and circumferences (+~1-3 cm, (2, 7, 8)). For circumferences the authors attributed the difference to anthropometrists compressing the skin by applying tension to the measuring tape. The study that used stationary cylinders reported overestimation of girth (0.7-0.9 cm) that was consistent for cylinders of different sizes (5). In our research we found similar levels of overestimation of height and HC, but AutoAnthro underestimated MUAC and underestimation only occurred among children with small arm circumferences. We reached the same overall conclusion on accuracy as authors from previous research — the finding of systematic inaccuracy does not diminish the potential of 3D imaging for anthropometry because the processing software can be adjusted to remove systematic bias. One caveat to our conclusion is that we recommended additional research before making MUAC adjustments, which we elaborate on in the following section.

For public health the most important factor when evaluating a 3D imaging system may be the ability to correctly classify nutritional status. Average bias provides little information on classification because it does not take into account the spread of differences. The International Organization for Standardization produced accuracy requirements for 3D imaging based on the confidence interval around average bias (5). In our research we measured classification agreement directly by looking at prevalence, sensitivity, and specificity; and found good agreement at the individual and population levels after adjusting to remove systematic inaccuracy. Classification agreement was presented in Chapter Four, and additional data are available in Figure 7-1. We did not identify any other studies that considered the ability of 3D imaging to correctly classify individuals by indicators of nutritional status.

7.3. Study Strengths and Limitations

Strengths and limitations for the individual studies that make up this research were covered in chapters three through six. Overall, the strengths of our research were that it was relevant because it responded directly to a call from the global nutrition community for the use of technology to improve anthropometric data quality and anthropometrists helped in the design of the research; and it was meaningful because scan-derived measurements were compared to gold-standard manual measurements, and anthropometrists helped to analyze and interpret data. Reported as limitations in previous chapters, there were study characteristics that need to be taken into account when interpreting or generalizing findings. In summary, children in our sample were well-nourished and of a known age, the sample was not representative of any geographic area, our anthropometrists were highly educated, processing scans was not fully automated, data was collected inside of large facilities, both anthropometrists and the study population were familiar with electronic devices, qualitative research was based on a small sample of anthropometrists, and qualitative data on caregivers and children was not direct. A full explanation of these limitations was provided in previous chapters and we expand on the implications of some of the limitations in the next section on research and development needs. One study limitation that was not fully described in previous chapters was that BST had access to manual measurement data before processing scans into measurements because all data was uploaded to a server that BST managed. In BINA we tested calibrated software on a new sample of children that were not involved in calibration, BST remained blinded to manual measurement data, and BST did not use manual measurement data from the new sample to process scans. Nonetheless, the lack of a firewall between the company and the manual measurement data leaves the study open to critique. In future studies the company should not have access to manual measurement data until after scan-derived measurements are produced.

7.4. AutoAnthro Research and Development Needs

In previous chapters we described a number of research needs to follow-up on BINA findings and to address study limitations, including the need to corroborate BINA reliability findings, to validate adjustments to remove systematic inaccuracy, and to determine if 3D imaging improves anthropometric data quality. An additional validation study can follow the same basic design of combining a comparison of scan derived measurements against gold, standard manual anthropometry with qualitative research that assesses the feasibility of using the imaging system in regular nutritional assessment, but a few modifications to the BINA study design are needed to address BINA limitations.

Like BINA, if another study compares scanning to high quality anthropometry, there will be no conclusions on the ability of 3D imaging to improve anthropometric data quality. A separate study that includes 3D imaging in a setting of poor quality manual anthropometry could provide evidence on improving data quality, or a validation study could be designed with a secondary objective of assessing the ability of 3D imaging to improve data quality. One possibility for assessing the ability of AutoAnthro to improve quality within a validation study is to incorporate anthropometrist training into a future study analysis plan. If inexperienced anthropometrists are used, a future validation study could evaluate the potential for anthropometric data quality improvement by comparing poor quality manual anthropometry collected during the training period to scan derived measurements carried out at the same time. In BINA we typically scanned children at the same location within a facility over a number of days. We experienced the most ‘software glitches’ attributed to lighting when moving from room to room in a hospital setting. A future validation study can be carried out in the context of a household survey or during community-based screening to assess scanner performance in different lighting conditions. The use of a household survey could also address another limitation of BINA, which is that the sample was not representative of the target age group. In BINA newborns and children under two years of age were overrepresented, and there were few children between one week and six months of age. Our findings on reliability and accuracy cannot be extrapolated to an entire age group, such as children under two years of age or children 0-5 years of age. A future validation study that includes a representative age structure is needed to draw conclusions on the appropriateness of the technology for an entire age group. A future validation study can also be carried out in a low-income setting to address additional BINA limitations. In BINA our sample was well-nourished; and we did not have a sufficient number of children with abnormal nutritional status to analyze prevalence or sensitivity/specificity for clinically significant indicators, such as obesity, wasting and severe stunting. Additional studies in populations with high prevalence of abnormal nutritional status are needed for appropriate evaluation of classification bias. Children and caregivers that participated in BINA were familiar with cameras and mobile devices. Our qualitative findings on efficiency, invasiveness and the user experience cannot be extrapolated to a setting where people are not familiar with technology; and there may be additional barriers to using the technology in a different population. During BINA recruitment caregivers often asked about potential harm from the scanner and privacy. Some caregivers did not consent for their children to participate in the study because of these concerns, but for the most part we were able to allay any fears by showing caregivers that the scanner image is not identifiable and explaining that the laser used in the 3D scanner poses no risk and that the same type of scanner is used by children in millions of households around the country in video game consoles. If carried out in a setting where people are not familiar with technology or have different views of technology, the qualitative component of a future validation study could gain valuable insight into additional barriers to the use of AutoAnthro.

In the remainder of this section we elaborate on our recommendations for future research and development that pertain to improving the quality of scan-derived anthropometry.

7.4.1. Research and development to improve the quality of scan-derived anthropometric data

In chapter four we found that reliability of scan-derived measurements was within 1 mm of gold-standard manual measurement reliability when using repeated scans (average of two sessions), which is a negligible difference and means that scan-derived anthropometry can be as reliable as manual measurement. Reliability of a single scan was slightly worse than manual measurement, but it was not clear if the decreased reliability made any meaningful difference. We concluded that additional research is needed to determine if the protocol for deriving measurements from 3D imaging should be changed from using a single-scan to repeated-scan, and stipulated that repeating scans would only add ~one minute to measurement time. The time motion study in chapter six gave a precise estimate for the amount of time that would be added by repeating scans — 68 seconds. Future studies or surveys using AutoAnthro should include at least two scan sessions for each child (resulting in at least two sets of measurements per child) until there is definitive evidence that a single session does or does not provide adequate reliability.

In chapter four we showed that scan-derived anthropometry was systematically inaccurate, and that inaccuracy led to misclassification of nutritional status. We recommended adjusting the software to remove systematic inaccuracy, and highlighted the need to carry out another study to see if our accuracy findings are repeated (or that the adjustments completely removed systematic inaccuracy). Adjustments to remove systematic inaccuracy can be made based on our average bias findings, but average bias figures should be corroborated by an analysis of why scanning produced inaccurate measurements. Also, it is important to note that we did not remove outliers in our analysis of average bias, and it may be necessary to remove outliers to determine the correct adjustment. Determining the source of inaccuracy should provide more confidence that adjustments are correct. In chapter five we speculated on the possible causes of inaccuracy for each measure; for stature and HC we hypothesized that the processing software did not adequately take into account the protocol for manual measurement. The BINA database is ideal for testing the stature and HC hypotheses because we are confident that the manual measurements were good quality. For MUAC the adjustment to remove systematic inaccuracy is more complex because accuracy was not consistent for children of different sizes, and we hypothesized that inaccuracy may have been a result of human/manual equipment error as opposed to scanning hardware or software error, speculating that with younger children the measuring tape was not completely flush to the skin around the entire arm circumference. In BINA we did not use the same measuring tapes that were used in the MGRS because they were not available from our manual equipment supplier. The measuring tapes used in MGRS were made of metal and may have been less susceptible to gapping. In chapters four and five we included children less than six months of age when analyzing MUAC accuracy, but MUAC is typically used for children 6-59 months of age. Additional analysis confirmed that scanning systematically underestimated MUAC for children 6-59 months of age and the bias was not consistent by size within this age group (Figure 7-1). Similar findings in the 6-59 month age group showed that more work is needed to determine the cause of inaccuracy. AutoAnthro should be tested on stationary cylinders of a known size, with the sizes corresponding to the MUAC sizes found in BINA, to confirm that systematic inaccuracy was not a result of hardware or software error. Future studies on AutoAnthro could use the same metal measuring tapes that were used in MGRS to determine if gapping was the cause of systematic inaccuracy.

7.4.2. Software development for automated scan processing

We altered the study protocol in response to study implementation challenges. First, at the beginning of data collection we shifted from taking only six scans per session to encouraging anthropometrists to take excess scans and select the six best scans for each session. That pivot was made to account for the difficulty of capturing a good quality scan when children were moving. Second, we allowed anthropometrists to alter the order of scans. Initially the protocol called for taking three front scans and then three back scans. During data collection anthropometrists found this impractical because they had to be opportunistic and take scans when the child was still, regardless of which direction the child was facing. After finishing data collection we attempted to adjust the processing software to automatically detect orientation (front/back), but it was not possible in the short timeframe available. We altered study protocol from complete automation of scan processing to manually coding child orientation in each scan. In its current state AutoAnthro requires manual input for both scan quality selection and child orientation.

For orientation we do not think it will be practical to require a specific scanning order in future data collection. Ideally, the processing software would be adjusted to automatically detect orientation. If automatic detection is not possible, an alternative would be to adjust the data collection component of the software to allow the anthropometrist to code orientation after taking each scan. The drawback to the latter approach is that it will increase measurement time, not only because selecting orientation will take time, but also because the increased time spent in between scans will make it more difficult to take advantage of the child being still or holding a pose.

When scanning children under five it is inevitable that movement will result in some poor quality scans, which means taking excess scans will continue to be necessary in future studies and software needs to be updated. To account for excess scans an immediate step to improving the software would be to adjust the data collection component of the software to allow the anthropometrist to mark a poor quality scan during data collection based on visual inspection, and to display a running tab of the number of 'not poor quality' scans completed. Software adjustment will facilitate scan selection and will help anthropometrists keep track of the number of completed scans, but it is not the ideal solution. The quality of the scan may not always be obvious to the anthropometrist, and like orientation selection, increased measurement time is a drawback of manually marking poor quality scans.

Experience during BINA illustrates the value of taking scans in quick succession. During data collection anthropometrists had the ability to manipulate the amount of time that they could view a captured scan. A longer viewing time allowed a more thorough visual inspection of the quality of the captured scan, but anthropometrists kept viewing time to a minimum so as not to lose the opportunity to finish all of their scans when a child was being cooperative. Software adjustment to mark scans of poor quality and select orientation should be considered a temporary solution because it does not automate scan quality control and it will not give anthropometrists more confidence that they are taking scans of an adequate quality; additional research is needed on scan quality to develop a more permanent solution, which is covered in the next section.

7.4.3. Scan quality research and development

The quality of 3D scans is dependent on the scanner, the software, the environment; and interactions between the three. Quality assessments of 3D images can look at noise (smoothness), outliers (data points that are out of place), and under-sampling (holes or missing pieces of subject) to quantify scan quality (9). In BINA we were able to control for outliers and under-sampling. Including six scans in a session enabled the processing software to fill in under-sampled areas in one scan with data from another scan in the final model. In addition, anthropometrists considered under-sampling during visual scan inspection to select the best quality scans, and the scan processing software automatically removed outliers. The BINA manual included scan quality criteria based on visual inspection and BST provided feedback on the quality of scans in pretesting and during the first month of data collection. Despite quality control efforts in BINA, at the end of data collection anthropometrists felt that frequently their scans of uncooperative children were not good quality. The perception of settling for 'subpar' scans, along with the difficulty in fostering cooperation of children six months to three years of age, led anthropometrists to conclude that AutoAnthro could not yet replace manual equipment, particularly for uncooperative children.

If anthropometrists often captured poor quality scans within a specific age group we could expect to see differences in measurement quality by age. In chapters four and five we found no differences in the accuracy or reliability of scan-derived measurements by age. There are a couple possible explanations for why quantitative results did not support the perception of settling for subpar scans. First, it is difficult to assess scan quality because a single measurement comes from processing and merging six scans. It is possible that anthropometrists did often settle for subpar scans, but that it was rare for an entire session (six scans) to be poor quality. Considering anthropometrists took excess scans it is possible they nearly always captured one or two scans of adequate quality for processing into measurements. A second explanation for the difference between perception and quantitative findings stems from our lack of knowledge of how much scan noise is acceptable for processing into accurate measurements. Noise, or how 3D points in close proximity relate to each other with respect to depth, can be thought of as a measure of surface smoothness. Processing software uses denoising algorithms to remove noise (10), but the underlying noise likely affects the quality of scan-derived anthropometric data. Motion and distance from the subject are two factors affecting scan noise, with the former being a highly variable factor when scanning young children. With visual inspection it was simple to detect a very smooth or very rough surface, but BINA anthropometrists had little confidence in visually judging the quality of scans that were somewhere in between smooth and rough, which was a common occurrence. BINA supervisors were not able to provide adequate instruction on assessing scan noise because we did not yet have data linking scan noise to measurement quality. The second explanation for the discrepancy is that anthropometrists may have judged scans that were not smooth as subpar, when in fact they were adequate for processing into accurate measurements.

We did not identify any studies that analyzed the effect of scan quality on anthropometric data. With BINA data we can now create a dataset that links scan quality to anthropometric data quality. Research on the associations between scan quality metrics and anthropometric data quality is needed to improve training and supervision, and to automate scan quality control. In BINA we used multiple metrics, such as reliability, accuracy and biological plausibility, to monitor the quality of manual measurements; but for scanning we relied solely on a simple list of visual inspection criteria. A first step for additional research on scan quality control is to evaluate the scans of children who were outliers in terms of scan-derived measurement reliability and accuracy. We did find a few outliers in the data and they may represent the rare case when scans were all poor quality. Displaying the scans of a session that resulted in inaccurate or unreliable measurement can help to train anthropometrists and will give supervisors a tool for monitoring and supervising scanning. The BINA dataset also provides the opportunity to carry out research that can facilitate automated scan quality control.

An ideal solution to scan quality control would be to adjust software to automatically predict the accuracy of scan-derived measurements from an entire session. 3D scan data was converted to anthropometry by fitting merged data from six scans to a scaled articulated model. There is a metric for how well the merged scan data fits the articulated model, referred to as model fit. Theoretically, model fit should be highly correlated with the accuracy of scan-derived anthropometry, and software could automatically detect the quality of a scan session based on model fit to inform the anthropometrist when more scans are needed for accurate measurement. However, model fit scores are not available until scans are processed. In BINA scans were uploaded to the cloud and processed after data collection was finished. It is not yet possible to immediately process scans on the local device (the tablet or phone that the scanner is connected to). It is possible to process scans on the local device during data collection, but in the near term processing will take at least five minutes. For most applications five minutes is probably too long to wait during data collection for feedback on the necessity of taking additional scans. Nonetheless, the association between model fit and measurement accuracy should be carried out now because it will be useful for supportive supervision and the processing time may be reduced enough by technology development to allow for immediate feedback.

Until model fit is immediately available, an alternative solution to predicting the accuracy of scan-derived measurements during data collection is needed. Research should be carried out on the correlations between anthropometric data quality and scan quality metrics that can be available immediately. A substantial amount of exploratory work is required for this research. We suspect noise was an important factor in BINA because of its relationship with movement and the lack of quality control. The previously mentioned analysis of scan sessions for production of training and supervision tools could help to determine which scan quality metric is most relevant to poor quality scan-derived anthropometric data. In addition, more exploratory work will be needed because all of the scan quality metrics that are immediately available describe a single scan. Research will need to determine how best to group the scan quality data when predicting a single measurement from six scans. For example, we do not know if average noise across all six scans or a specific level of noise among fewer scans is the best predictor of anthropometric data quality. Research will also need to determine what combination of scan quality metrics is the best predictor: noise alone, noise and distance from the subject, noise and under-sampling, etc.

It is worth noting that research on scan quality using the BINA database is limited because we did not keep all of the scans captured during the study. Anthropometrists deleted excess scans during data collection because we did not have sufficient server space, and so the BINA database does not contain most of the poor quality scans. With BINA data it is not possible to determine how often scans are of insufficient quality for processing into correct measurements, and we cannot know how much cooperation affects scan-derived measurement quality when anthropometrists take only six scans. It is possible that quality is rarely affected by movement with the current software, and if that is the case, it may be sufficient in a research setting to offer improved supervision that gives anthropometrists confidence that they are able to control quality and are collecting scans of adequate quality. Nonetheless, the pursuit of automated scan quality control is worthwhile, even if measurement quality is rarely affected, because automation will lead to more uniform results from different anthropometrists and could facilitate the regular use of 3D imaging for child anthropometry in clinics and surveys.

7.5. 3D Imaging for Anthropometry – The Future

7.5.1. Role in improving the use of anthropometric data

In the background we described an attempt by the Body Benchmark Study to bring 3D imaging to adult nutritional assessment through the development of a new anthropometric indicator. The development of AutoAnthro took a different approach, responding to a call to improve the quality of current measurements, specifically length. BINA was not designed to evaluate if AutoAnthro improved anthropometric data quality, but it was a first step in evaluating a potential technological solution to poor quality anthropometry. We were not able to draw any conclusions on improving data quality because our anthropometrists produced high quality measurements with both 3D imaging and manual measurement. It is encouraging that scanning produced reliable measurements despite spending substantially less time on training and supervision for 3D scanning than manual measurement, and that scan measurement reliability was not affected by age or the anthropometrist; but these findings did not provide conclusive evidence on quality improvement. When considering the future potential of 3D imaging to improve anthropometric data quality it is important to consider that most child anthropometric indicators coming from large scale surveys also include age, and in many countries misreported age is a common source of error (11). From 2000 to 2010 the global average of birth registration increased from 58% to 65% (12), and as birth registration improves age misreporting will be less of an issue, but presently in least developed countries only 38% of births are registered (12). New technology for measurement will need to be combined with efforts to accurately assess age to fully improve anthropometric data quality in many parts of the world in the near term.

The call for technology came from global nutrition and we designed our research through that lens. We focused primarily on measurements that are used as indicators to monitor infant and young child underdevelopment. These indicators are relevant to any setting, but in places that are experiencing the obesity epidemic the measurements that are used as proxies for body composition are also important, and they too suffer from poor data quality. In the US measurements obtained from manual anthropometry (without calculating volume) are collected in national surveys and used as proxies for body composition, but they have not yet supplanted BMI in regular nutritional assessment. In the US in 2016 the prevalence of overweight and obesity were determined using weight-for-height (WHZ) for children under 2 and body mass index (BMI) for children over 2 (13, 14). WHZ and BMI are proxies for total body fat that do not directly differentiate between fat and fat free mass (FM and FFM). In addition, WHZ and BMI do not provide information on body shape and cannot distinguish between subcutaneous and visceral fat. Since abdominal visceral fat is strongly associated with metabolic disturbances and disease, it is useful to have measures of both total body fat and the fat distribution (15). Waist circumference (WC) and sagittal abdominal diameter (SAD) are proxies for abdominal visceral adiposity. In 2016 the US National Health and Nutrition Examination Survey (NHANES) included waist circumference (WC) for children over 2 years of age and adults (16). A World Health Organization (WHO) expert committee recommended the use of waist circumference alongside BMI (17) and the US Centers for Disease Control and Prevention (US CDC) has reference values for waist circumference (18). The waist circumference to height ratio (WHtR) is considered a better predictor of morbidity risk compared to BMI for adults (19), WHO and multiple reviews advocated for the use of WHtR alongside BMI for children because of improved risk prediction from using multiple measures (17, 20, 21) Sagittal abdominal diameter is also considered to be a better measure of risk than BMI (16). Despite a substantial amount of evidence and advocacy for the use of WC, the measurement is not common in regular nutritional screening because it is not easy to measure and is often unreliable. In addition to improving the quality of measurements commonly used in developing country surveys (length), 3D imaging could also help to improve the quality of measurements used to monitor the obesity epidemic. By improving quality and making the measurement easier, 3D imaging could facilitate the use of waist circumference in regular nutritional assessment, and as the obesity epidemic spreads around the globe obesity related measurements derived from 3D scanning would likely become a regular part of nutritional surveys and assessment in all countries.

We carried out this research on apparently healthy children. In presenting our research design and findings to clinicians it became obvious that while they were interested in our results, they also wanted to know if 3D imaging for anthropometry could work for patients who cannot easily move into the positions required for manual measurement. There was specific interest in measuring patients with cerebral palsy and those that are bed-ridden, and clinicians were interested in estimating weight with the scanner. The AutoAnthro scan processing software did not require specific poses for children under two years of age, and the lack of posing did not seem to affect the quality of scan-derived measurements. If poses are not necessary, it is likely that AutoAnthro could also be used to measure surface morphology of people that are immobile. Also, the company that developed AutoAnthro has experience with estimating weight from 3D scans, and further technology development could make it possible to extract an accurate weight in the near future. 3D imaging could play a role in improving the regular nutritional assessment of all people, regardless of their mobility status. The use of a portable 3D scanner for a specific problem may be an efficient way to start using the technology in the health sector. In the background we described StarScanner, the only other 3D imaging system designed for children. StarScanner addressed one specific need in the health sector and it is now used in hospitals in many. Bringing a solution to the measurement of immobile people could familiarize medical staff with the technology and facilitate wider use of AutoAnthro in the health sector. Barriers to the use of 3D imaging in the health sector are discussed further in the next section.

7.5.2. Technology development, cost and bringing down entry barriers

Researchers and organizations working with 3D imaging for anthropometry need to consider that 3D scanners are being developed for a variety of purposes outside the health sector, and that the rapid pace of technology development may quickly affect recommendations from our research. We highlighted that AutoAnthro did not work in direct sunlight and that lighting may be an issue when using the technology in a household survey or community screening setting. Occipital, the same company that developed the scanner used in AutoAnthro, recently developed a new 3D scanner (Structure Core) that reportedly functions outside in direct sunlight (A. Rodnitzky, personal communication with product brief, September 12, 2017). Improvement of 3D imaging technology will continue, with scanners gaining new capabilities and producing better quality scans. In our discussion on accuracy and reliability of scan-derived anthropometry we covered the factors that affect scan quality with the current hardware, such as movement. There are two ways technology development could influence our recommendations for quality control. First, the precision of the scanner itself will likely change over time and more precise scans could improve the quality of scan-derived anthropometry. Second, the factors that affect scan quality of Structure Sensor may have no impact on the scan quality of future scanners.

In Chapter six we discussed the cost of AutoAnthro hardware and concluded that efficiency gains related to training, staff, and measurement time could offset the increased cost compared to manual equipment. In late 2017, as this dissertation was being finalized, Apple unveiled a mobile phone (iPhone X) that has light coding technology built into the phone, with the scanner branded as a TrueDepth camera and marketed as a tool for face recognition. The iPhone X retails at $999 (22); a year from now when a new model is released it could be half that amount. Stand-alone light coding scanners will likely drop their price to compete, and so it is reasonable to predict that in the very near future, within a year or two, a portable 3D scanner will cost no more than a length board; and it may only be a matter of time before a mobile phone with a 3D scanner is at a similar price. Our research can be used to inform a costing study on 3D imaging for anthropometry, but any research needs to take into account the declining cost of scanners, and it may soon be a moot point if a 3D scanner is cheaper than manual equipment. Another cost consideration that we did not previously discuss is related to the value of the collected anthropometric data. In current health and nutrition surveys we collect a few measurements that are of little interest to people outside of the health sector. In using 3D imaging to improve anthropometric data quality, health and nutrition surveys would also be collecting data that interests industry. National sizing surveys are supported by industry; 3D imaging will bring new opportunities for public private partnership in carrying out large scale health and nutrition surveys and research.

In the background we described the barriers to the use of 3D imaging for regular nutritional assessment in the health sector, namely cost and dedicated space. The cost barrier was likely removed in recent years, and if 3D scanners are not already affordable enough, they will be in the near future. Development of the portable 3D scanner used in our research could have also removed the barrier of dedicated space because 3D imaging no longer requires an entire room to be setup with cameras in fixed positions. AutoAnthro does have potential for regular use in the health sector, but there are characteristics of the imaging system that could hinder entry or affect long-term use — the person being scanned cannot be touched by anyone and has to be undressed to their undergarments. Privacy concerns related to undressing could be resolved in most contexts and might not be a significant barrier to the use of AutoAnthro in the health sector. Likewise, development of automated quality control may help to enable widespread use of an imaging system that does not allow touching. Even if these characteristics do not affect entry, it is worth considering how future technologies could affect the use of AutoAnthro. Past innovation in child anthropometric equipment may shed light on future developments. Prior to recent calls to replace length boards with new technology, salter scales were largely replaced by digital scales with a taring function. The digital scales helped to reduce invasiveness by promoting cooperation from the child, who could be held in their mother's arms, and reduced random error because of less child movement and less misreading of the scale by the measurer. Digital scales replaced salter scales not only because they were more accurate, but because they made the measuring experience easier and more pleasant for the child, the caregiver, and the anthropometrist. It is possible that a 3D imaging system that restricts touch and requires undressing will be supplanted by an imaging system that was designed for the mother to hold a child.

7.5.3. Moving toward Better Indicators and 3D Measurements

Previously, we described how the garment industry was an early supporter of 3D imaging for anthropometry because clothing design requires a large number of measurements. In the health sector anthropometry is primarily limited to a few measurements: weight and height for everyone, HC for newborns, WC for pregnant women, and MUAC for screening in low-resource settings. The use of multiple measurements has been stymied by the practical constraints of manual measurement. It appears that in recent years technology development has removed the main barriers to the regular use of 3D imaging in the health sector. The use of 3D imaging for anthropometry will provide an opportunity to make sure all of the known, important measurements are regularly captured and analyzed. One example is making waist circumference a part of regular nutritional assessment, which we discussed in a previous section. Another example is body fat. Body fat can be estimated from weight and body volume, but volume is extremely difficult to estimate with manual, 1D measurements. Currently, researchers use relatively expensive tests (air displacement plethysmography and dual-energy x-ray absorptiometry) to measure body volume, and the measure is not a part of regular nutritional assessment. Studies have shown that calculations of percent body fat based on scan derived measurements of body volume are both reliable and accurate (7, 23, 24). 3D imaging could help to make estimated body fat a routine measurement.

As 3D imaging for anthropometry becomes more common in the health sector there will be an opportunity to come up with novel measurements that are better predictors of outcomes of interest, but the development of such measures is dependent on longitudinal data that is connected to health outcomes. Some of the research on 3D imaging referenced in the background comes from this type of study, a large-scale, (n=10,000) longitudinal study on disease risk factors (25, 26). Another potential source of data is medical records. If scanning becomes a part of regular health checks, the data could be used to develop novel anthropometric indicators, especially if it is stored in a 3D format. Many of the new indicators that come from 3D imaging are likely to be based on 3D measures. In the background we described the development of two new indicators based on 3D measures, and our findings on head circumference also provide an illustrative example of the potential for 3D measures. The sensitivity and specificity of 3D imaging for HC was worse than stature and MUAC, and we attributed the difference to the lack of a fixed position for manual measurement of HC. With manual measurement we try to routinely find the largest circumference, but that can be difficult because everyone has a head that is shaped differently. With HC we measure one circumference of the cranium to represent the entire cranium. 3D imaging makes it feasible to routinely measure cranial volume, which is a direct measure of cranium size that is probably a better predictor than HC. If 3D imaging for anthropometry sees a large jump in usage, one of the challenges in facilitating the development of new indicators will be to create a mechanism to share and pool data.

7.5.4. Connecting BINA to future development of 3D imaging for anthropometry

Our research focused primarily on using 3D imaging for measurements that are currently used in regular nutritional assessment of children, with a focus on indicators used for the assessment of undernutrition. Also, we restricted our analysis to healthy children to extrapolate findings to the majority of children. However, with an eye to the future, we also included weight, measured waist circumference on 11 children, and measured one child with cerebral palsy. We discussed how previous research showed that 3D imaging was appropriate for measurements related to the obesity epidemic, but that research and efforts to mainstream novel indicators have not been successful to incorporate 3D imaging into regular nutritional assessment. We did not identify any research on the use of 3D imaging for anthropometry of immobile populations. AutoAnthro addressed the main barriers of using the technology in the health sector by using a low cost scanner that is portable; and extended the use of 3D imaging for anthropometry to children by developing software capable of handling movement. Additional research is needed to determine if AutoAnthro can be used for regular nutritional assessment of children in the context of the obesity epidemic and for nutritional assessment of immobile populations. An initial step could be to evaluate the use of AutoAnthro for indicators that are currently in use. Additional analysis of BINA data that uses weight can evaluate the accuracy and reliability of scan-derived child BMI. We discussed that waist circumference is currently recommended for nutritional assessment of children; BINA data from the 11 children with WC can be used to develop and calibrate AutoAnthro processing software to measure WC. BINA data and anthropometrist's experience scanning a child with cerebral palsy can be used as a proof of concept for measuring immobile populations, and further development of AutoAnthro processing software to measure body volume could provide another proof of concept — that AutoAnthro, like other 3D imaging systems, can be used in place of more expensive techniques to measure body fat. Ultimately, additional studies that evaluate scan-derived waist circumference and body volume using an inexpensive, handheld scanner with software developed for children is needed. Like in previous studies of other more expensive 3D imaging systems, scan derived waist circumference can be compared to gold-standard manual measurements and body volume can be compared to gold standard machine measurement (BodPod, DEXA, or MRI). Our findings along with additional analysis of BINA data can help to design such a study.

7.6. Conclusions

Manual measurement did provide high-quality anthropometry in this research and it is possible to improve manual measurement of circumferences and length in clinics and surveys. However, there may be limited institutional need and motivation to improve quality; and improved technology could be the most efficient driver of widespread quality improvement. We do not yet know if AutoAnthro will lead to improved quality of child anthropometric data, but BINA showed that a 3D imaging system produced reliable measurements of children under five years of age, which suggests that 3D imaging can be an appropriate anthropometric tool for infants and young children. Further research and development is needed, particularly to determine if AutoAnthro improves quality and to address our findings of systematic inaccuracy and anthropometrists' lack of confidence in scanning uncooperative children. The potential value of 3D imaging for anthropometry is not limited to quality improvement; adoption of the technology could result in higher use of a variety of anthropometric indicators in regular nutritional assessment, and the discovery of new measurements that make anthropometry a better predictor of outcomes of interest.

7.7 Supplementary Tables and Figures

Table 7-1 Sensitivity and specificity of adjusted, scan-derived measures when compared to best-estimate manual measures among children 1-59.9 months of age

	<-1 SD			**>+1 SD**		
	Prevalence	Sensitivity	Specificity	Prevalence	Sensitivity	Specificity
Stature						
GS Manual	21.9	.	.	11.0	.	.
Single Manual	22.7	0.95	0.98	11.0	0.95	0.99
Single Scan	23.5	0.92	0.96	12.0	0.91	0.98
Repeated Scan	22.7	0.93	0.97	10.7	0.91	0.99
Head Circumference						
GS Manual	9.4	.	.	27.3	.	.
Single Manual	9.7	0.92	0.99	27.3	0.94	0.98
Single Scan	9.9	0.78	0.97	26.8	0.84	0.95
Repeated Scan	9.2	0.81	0.98	26.5	0.87	0.96
Arm Circumference						
GS Manual	2.8	.	.	41.6	.	.
Single Manual	3.6	1	0.99	41.8	0.93	0.95
Single Scan	4.7	0.91	0.98	41.6	0.91	0.94
Repeated Scan	3.9	1	0.99	40.5	0.93	0.96

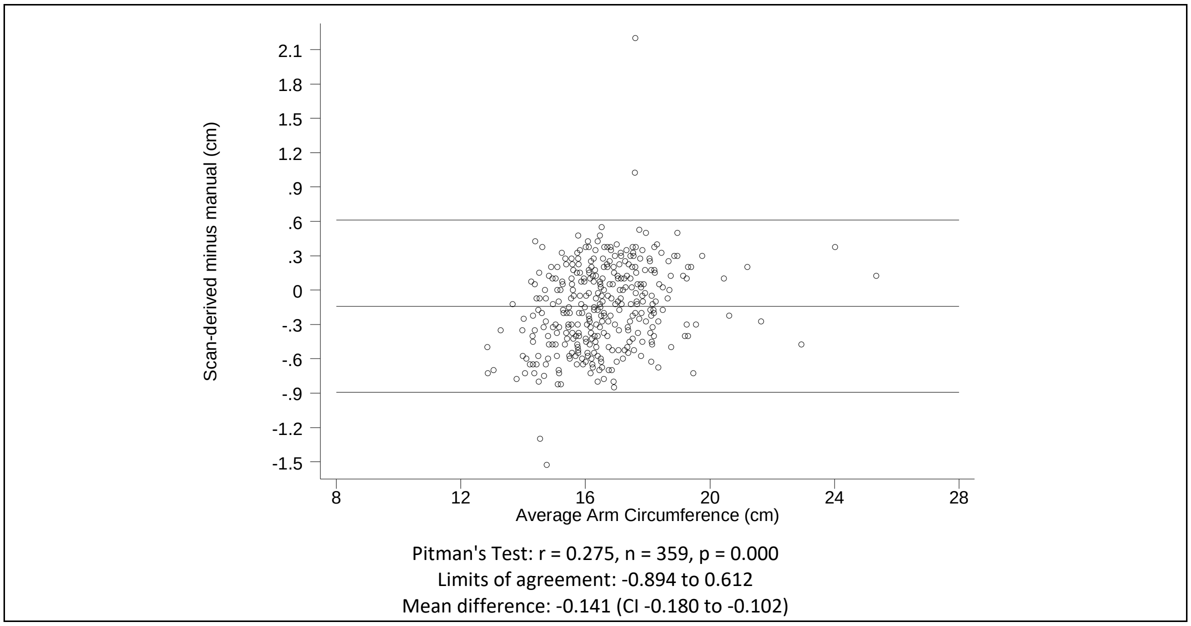


Figure 7-1 Bland-Altman plot of best-estimate manual measurements subtracted from single-scan for MUAC among children 6-59 months of age

References (Chapter 7)

1. Kuehnapfel A, Ahnert P, Loeffler M, Broda A, Scholz M. Reliability of 3D laser-based anthropometry and comparison with classical anthropometry. Scientific reports. 2016;6:26672.

2. 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;10(3):e0119430.

3. Cook T. Using the Microsoft Kinect for patient size estimation and radiation dose normalization: proof of concept and initial validation. 2013.

4. Soileau L, Bautista D, Johnson C, Gao C, Zhang K, Li X, et al. Automated anthropometric phenotyping with novel Kinect-based three-dimensional imaging method: comparison with a reference laser imaging system. Eur J Clin Nutr. 2016;70(4):475-81.

5. Clarkson S, Wheat J, Heller B, Choppin S. Assessment of a Microsoft Kinect-based 3D scanning system for taking body segment girth measurements: a comparison to ISAK and ISO standards. J Sports Sci. 2016;34(11):1006-14.

6. Pepper MR, Freeland-Graves JH, Yu W, Stanforth PR, Cahill JM, Mahometa M, et al. Validation of a 3-dimensional laser body scanner for assessment of waist and hip circumference. Journal of the American College of Nutrition. 2010;29(3):179-88.

7. Wang J, Gallagher D, Thornton JC, Yu W, Horlick M, Pi-Sunyer FX. Validation of a 3-dimensional photonic scanner for the measurement of body volumes, dimensions, and percentage body fat. Am J Clin Nutr. 2006;83(4):809-16.

8. Wells JC, Treleaven P, Cole TJ. BMI compared with 3-dimensional body shape: the UK National Sizing Survey. Am J Clin Nutr. 2007;85(2):419-25.

9. Weyrich T, Pauly M, Keiser R, Heinzle S, Scandella S, Gross M. Post-processing of scanned 3D surface data. Proceedings of the First Eurographics conference on Point-Based Graphics; Switzerland. 2386347: Eurographics Association; 2004. p. 85-94.

10. Sun X, Rosin PL, Martin RR, Langbein FC. Noise analysis and synthesis for 3D laser depth scanners. Graphical Models. 2009;71(2):34-48.

11. Grellety E, Golden MH. The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition. PloS one. 2016;11(12):e0168585.

12. United Nations Children's Fund. Every Child's Birth Right: Inequities and trends in birth registration. New York; 2013.

13. Fryar C, Carroll, M., Ogden, C. Prevalence of High Weight-for-recumbent Length Among Infants and Toddlers From Birth to 24 Months of Age: United States, 1971–1974 Through 2013–2014. In: Surveys DoHaNE, editor. Health E-Stats: National Center for Health Statistics; 2016.

14. Fryar C, Carroll, M., Ogden, C. Prevalence of Overweight and Obesity Among Children and Adolescents Aged 2–19 Years: United States, 1963–1965 Through 2013–

2014. In: Surveys DoHaNE, editor. Health E-Stats: National Center for Health Statistics; 2016.

15. Gibson RS. Principles of nutritional assessment. 2nd ed. New York: Oxford University Press; 2005. xx, 908 p. p.

16. US CDC. National Health and Nutrition Examination Survey Anthropometry Procedures Manual. Atlanta: US Centers for Disease Control and Prevention; 2016.

17. World Health Organization. Waist circumference and waist-hip ratio : report of a WHO expert consultation, Geneva, 8-11 December 2008. Geneva: World Health Organization; 2011. vi, 39 p. p.

18. National Center for Health Statistics (U.S.), National Health and Nutrition Examination Survey (U.S.). Anthropometric reference data for children and adults : United States, 2011-2014. Hyattsville, Maryland: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2016. p. p.

19. Madden AM, Smith S. Body composition and morphological assessment of nutritional status in adults: a review of anthropometric variables. J Hum Nutr Diet. 2016;29(1):7-25.

20. Horan M, Gibney E, Molloy E, McAuliffe F. Methodologies to assess paediatric adiposity. Ir J Med Sci. 2015;184(1):53-68.

21. Wells JC, Fewtrell MS. Measuring body composition. Arch Dis Child. 2006;91(7):612-7.

22. Pierce D. Meet the Iphone X, Apple's new high-end handset. Wired. 2017.

23. Garlie TN, Obusek JP, Corner BD, Zambraski EJ. Comparison of body fat estimates using 3D digital laser scans, direct manual anthropometry, and DXA in men. Am J Hum Biol. 2010;22(5):695-701.

24. Adler C, Steinbrecher A, Jaeschke L, Mahler A, Boschmann M, Jeran S, et al. Validity and reliability of total body volume and relative body fat mass from a 3-dimensional photonic body surface scanner. PloS one. 2017;12(7):e0180201.

25. Kuehnapfel A, Ahnert P, Loeffler M, Scholz M. Body surface assessment with 3D laser-based anthropometry: reliability, validation, and improvement of empirical surface formulae. Eur J Appl Physiol. 2017;117(2):371-80.

26. Loffler-Wirth H, Willscher E, Ahnert P, Wirkner K, Engel C, Loeffler M, et al. Novel Anthropometry Based on 3D-Bodyscans Applied to a Large Population Based Cohort. PloS one. 2016;11(7):e0159887.