The Pennsylvania State University
The Graduate School

REWEIGHTING ANTHROPOMETRIC DATA USING A NEAREST NEIGHBOR APPROACH

A Thesis in
Engineering Design
by
Kannan Anil Kumar

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2017

The thesis of Kannan Anil Kumar was reviewed and approved* by the following:

Matthew B. Parkinson
Associate Professor of Engineering Design
Thesis Advisor

Scarlett Miller
Assistant Professor of Engineering Design

Sven G. Bilén
Associate Professor of Engineering Design,
Head of School of Engineering Design, Technology and Professional Programs

*Signatures are on file in the Graduate School.

Abstract

A new method to reweight anthropometric data from a reference population to match that of a target population is proposed. When designing products and environments, detailed data on body size and shape are seldom available for the specific user population. Instead, the available data are outdated or represent a population that is demographically different on factors that are known to affect anthropometry. One way to mitigate this issue is to *reweight* available data such that they provide an accurate estimate of the target population of interest. This is done by assigning a statistical weight to each individual in the reference data, increasing or decreasing their influence on statistical models of the whole. This paper presents a new approach to reweighting these data. Instead of stratified sampling (the traditional approach), the proposed method uses a clustering algorithm to identify relationships between the target and reference populations using their height, mass, and body mass index. The newly weighted data were shown to provide more accurate estimates than traditional approaches. Data weighted with the new approach was used in different multivariate design test cases to demonstrate its use in real-world design applications. The improved accuracy that accompanies this method provides designers with an alternative to data synthesis techniques as they seek appropriate data to guide their design practice.

Table of Contents

List of Figures

List of Tables

List of Symbols

b Bin number, p. 14

n_w Record number in the weighted data, p. 14

S_1 Stature of a record from the set of data to be reweighted, p. 15

S_2 Stature of a record from the weighted set of data, p. 15

M_1 Mass of a record from the set of data to be weighted, p. 15

M_2 Mass of a record from the weighted set of data, p. 15

B_1 BMI of a record from the set of data to be weighted, p. 15

B_2 BMI of a record from the weighted set of data, p. 15

W_0 Weight to be assigned to each of the remaining records in a bin, p. 16

W_w Weight to be assigned to each of the weighted records in a bin, p. 16

N_0 Number of records from the data to be weighted which are assigned the same bin number as the bin under consideration , p. 16

Acknowledgments

I would like to acknowledge the contributions of those who worked on aspects of this problem before me. I would also like to acknowledge the work of my thesis advisor, Dr. Matthew Parkinson for his work on human anthropometric data and design for human variability, and thank him for his guidance and assistance throughout my graduate endeavors.

Dedication

I would like to thank my family for their support in innumerable ways for me. I also thank my colleagues in Open Design Lab for their assistance, and the friends I made in Pennsylvania State University for their support.

Chapter 1

Intoduction

Anthropometric data quantify the body size and shape of individuals. They are used extensively for the ergonomic design of artifacts, tasks, and environments since body dimensions play a critical role in their design. Factors such as product dimensions, shape, and other physical parameters influence the fit, convenience, comfort, and performance of the designed item. Guidance on how to conduct these design and assessment activities is provided through ergonomics curricula [2] and textbooks [3, 4], discipline-specific guides (e.g., [5]) and standards [6, 7, 8]. Broadly speaking, a person is classified as *accommodated* if the design can be used safely and comfortably in the intended manner. This means that every body measure relevant to the interaction of the human body with the product must be accommodated by the corresponding product dimension. For example in the typical case of a chair, height of the chair must be less than or equal to the popliteal height of the user, width of the chair must be greater than or equal to sitting hip breadth of the user and so on. A person is considered accommodated in this chair only if all these relevant body measures are accommodated by the corresponding dimensions of the chair at the same time. The conventional method of identifying the spatial requirements of a user population requires the identification of values that represent the extremes (e.g., the 5^{th} or 95^{th} percentiles) of variability that might be observed in a measure of interest [9]. Depending on the individual design variable and overall accommodation targets, the required value might be a minimum, maximum, or range. For example, doorway height might have an established maximum and the vertical adjustability of a task chair would have minimum and maximum values. These might be determined through available anthropometric data and the desired levels of accommodation.

Of course this approach relies on accurate estimates of the body size and shape of the target user population. Since data for the precise user population are seldom available, designers must assume that available data are "close enough". However, several determining factors including the demographic factors such as nationality, race, age distribution, and gender ratio cause the anthropometric measures to vary between different populations. Different occupational groups are also observed to have distinctive anthropometric characteristics which differ greatly from the general population [10]. Additionally, many of the databases of anthropometry are decades old. The 1988 U.S. Army anthropometric survey (ANSUR), for example, was collected in the mid 1980s [11]. Nevertheless, it is widely used in design due to its easy availability and quality and volume of detailed measurements. However, the race, age, and fitness of this sample are very different from those of the general US population. Additionally, changes are observed in population characteristics in the time since 1988 due to secular and demographics trends [12]. These changes have also occurred within military populations. The increase in several anthropometric measurements of active U.S. Army from 1988 to 2007 was demonstrated through a pilot study [13].

Therefore, many products are designed for the wrong population due to unavailability of detailed anthropometric data for their target populations. This causes the accommodation of the designed product to be considerably less than what was estimated, thereby only allowing a lot less number of users to use the product conveniently. Such disaccommodation can lead to a variety of health, safety, performance, and functionality issues are described in section 2.1. This thesis presents a new methodology to reweight detailed anthropometric data using stature, mass and BMI to match the target population. A cluster analysis is performed by measuring the closeness of the predictor values based on their euclidean distance. The data points are clustered together based on their characteristic(s) so that the same statistical weight can be assigned to these data points. Further binning can be performed using factors such as age, gender, and race. This allows us to obtain detailed anthropometric data for the target population when only two of the three measures, stature, mass, and BMI, are known for the target population. It is also easier to survey two such measures compared to the time, effort and cost associated with a detailed anthropometric survey.

1.1 Research Objective

The motivation for this research is the mismatch between the designed products and the populations that they were intended for, resulting in far less people being able to use the products than intended. This occurs due to differences between the design population and the target population of a design. Most products are designed using ANSUR data (refer section 2.2.1) collected from US army population in the 1980s, due to its easy and free availability. So for example, when a car is designed to accommodate 90% of the U.S. civilian population using this data, in reality a far lower fraction of the U.S. civilian population would be able to use the car comfortably and safely, due to the wider variety in the body shape and size of civilians compared to army personnel. Unavailability of detailed anthropometric data for the target population of a product to be designed. This forces the designers to use the available data instead. The resulting designs would therefore accommodate people with different maximum or minimum anthropometric measurements. This results in either reduced accommodation which would affect the market for the product, or increased accommodation, in which the cost of the product would have unnecessarily been higher.

The objective of this research is to develop a new approach to reweight data so to represent a target population, and compare the accuracy of this method with those of the existing ones.

The next chapter looks at the existing literature in the field. Three different methods used previously to reweight data are explained. In chapter 3, these methods are then used to reweight anthropometric data to match a target population, and their results are discussed. Then the new method is introduced and explained. It is followed by experimenting with the new method to identify the effects of different parameters on the accuracy of the reweighted data. Several ways to validate the accuracy of reweighted data by comparing with the original representative data are also demonstrated in this process. In chapter 4, this new method to reweight data is used to reweight a particular dataset to represent different populations. The results of these reweighting trials are discussed along with the reasons for errors. Finally in chapter 5, three design test cases are used to demonstrate the use of the new methods in real life design situations. The accommodation level of these designs were found using the actual data and the reweighted data, and compared.

Chapter 2

Literature Review

This chapter introduces the relevance of proper sizing of products and the health, safety, performance and functionality issues associated with their improper sizing. Different methods to obtain anthropometric data for the target population are then described. The different commonly used anthropometric databases and their drawbacks in design purposes are further discussed. Finally the existing reweighting methods to obtain anthropometric data is used to reweight data to observe the level of errors, thereby demonstrating the requirement for a new method.

2.1 Issues arising from improper sizing of products and workspaces

Improper sizing and seat positioning in cars have shown to increase the response time and therefore the braking distance while driving[14]. The distance between the headform and the head restraint in car seats affects the occurrence of whiplash-associated disorders caused by rear impacts[15]. These cause major safety concerns with improperly sized seating and vehicle packaging. Improper sizing of seats and workspace could cause users to sit in wrong postures for extended periods of time causing several health issues. This is particularly an issue in case of long and frequent driving scenarios such as in case of truck drivers. Alsong with its safey and health issues, the performance of the user many also be lowered.

Workspace configuration design has been suggested to be important in head stabilization of ship operators[16], where the head movement of the operator has to be minimized in order to reduce motion sickness[17] and overall seasickness among officers of the U.S. Coast Guard. The objective of this study was to improve the seakeeping by the U.S. Coast Guard. Similar restriction of head movement has also been found to reduce airsickness among aircraft pilots[18, 19]. In case of vehicles and particularly aircraft, proper vision of the surroundings and access to all the controls are important, which can be ensured by designing with the relevant anthropometric data. It was found in a study that designing airplanes using preestablished critical limits for different dimensions caused 52.6% of the target naval aviators to be disaccommodated by the airplane cockpit design as opposed to the intended 10% disaccommodation[20]. This unintended disaccommodation could be avoided if the design was virtually tested with the anthropometric data of the target pilot population.

It is seen that the anthropometry of people of different occupational groups vary greatly[10]. Therefore relevant anthropometric databases must be chosen to design workspace and personal protective equipments for use in a manufacturing environment in order to ensure the safety and productivity of the workers. Anthropometry may be used in workspace design in order to ensure sufficient clearances from the body to surrounding hazards such as the equipment itself, to ensure that there is no hindrance to the operator's movements by any parts of the machine in the workspace, and to calculate the sufficient distances from and between the controls. Non-ergonomic design of tools have been found to cause health issues. A study on workers in the carpet weaving industry concluded that constant and repetitive tying of knots with non-ergonomic hand tools could result in arthritis, neuralgia, and permanent deformation of fingers [21]. Such issues can be countered up to a certain extent by proper sizing of the designs which requires appropriate anthropometric data. But in cases where detailed anthropometric data are not available for the target population, the resulting designs would still pose the same issues and risks. Appropriate anthropometric data and digital human models also help study occupational biomechanics which in turn assist workspace design, thereby reducing musculoskeletal injuries and disabilities in industries[22, 3, 23].

2.2 Anthropometric data used

Several different anthropometric databases are available with varying levels of detail, aimed at populations different in terms of nationality, occupation, etc. Four anthropometric databases are used in this thesis for the different studies conducted. These anthropometric databases, their advantages, and their shortcomings are discussed in the following sections.

2.2.1 ANSUR

One of the first detailed anthropometric database is the 1988 Anthropometric Survey of U.S. Army Personnel, commonly known as ANSUR[11]. This survey was conducted from 1987 to 1988 among the Army personnel in the U.S. It was then downsampled from nearly 9000 men and women to create a subset of 1774 men and 2208 women that matched the age, and racial composition of the US army in active duty in June 1988. It contained 132 standard measurements, 60 derived dimensions which were calculated from the standard measurements, and 48 head and face dimensions. This data is used in the design of product for civilians even in other countries owing to its high level of detail, and free and easy access. But this would result in products designed with the required dimensions only when the target population matches that of the U.S. army population in 1988. Army personnel tend to have better fitness compared to most of the civilian populations. Apart from the generic differences between army and civilian populations, the ANSUR data has been found to not represent the modern U.S. Army population accurately, due to secular trends[24]. These secular trends include steady changes particularly in stature and BMI over a period of several years.

2.2.2 NHANES

The National Health and Nutrition Examination Survey (NHANES) is a program of studies that aimed at assessing the health and nutritional status of the U.S. population[25]. This involved conducting a demographics survey of the people along with several other surveys relevant to health and nutrition. The NHANES data is representative of the U.S. population and is updated every two years. For this reason, trends in the U.S. population over the year can be studied with these data. The data released every two years can also be combined for a larger sample size while still being a representative sample. The major shortcoming of this data is that useful anthropometric data is mostly limited to stature, mass, and BMI along with demographic information such as gender, age, and race. Although detailed anthropometric data are not available, NHANES is an ideal candidate to reweight other detailed anthropometric data to represent the current civilian U.S. population.

2.2.3 CAESAR

Civilian American and European Surface Anthropometry Resource (CAESAR) surveyed the civilian populations of three countries which represent the North Alantic Treaty Organization (NATO)[26]. These countries included the U.S.A., the Netherlands and Italy[27, 28]. Due to the inclusion of a site in Canada in the U.S.A. sample, it is referred to as the North American CAESAR sample. It surveyed civilians aged 18-65 in these countries with 3-D measurement technology. It was a collaborative effort of many industries and partners in different countries. The CAESAR data contains detailed anthropometric data derived from the 3D scans. The body scan data allows to make more detailed digital human models for design purposes. But a major drawback of this data is that it is not representative of any population. It would be inappropriate to use the CAESAR data as such in design as the products would then not be designed for any particular population. Therefore it has to be weighted to represent the target population before its use for most design purposes.

2.2.4 AIST

The AIST data contains detailed anthropometric data with 255 different body dimensions collected in Japan from 1991 to 1992[29]. It consists of two age groups, the civilian Japanese youth aged 18 to 29 years and the older Japanese civilians ages 60 years and older. The former contains 200 females and 200 males, while the latter contains 50 females and 50 females. This population will herein be referred to as Japanese population in this thesis. This data is assumed to represent the population in Japan and therefore other detailed anthropometric data is reweighted using the AIST data in this thesis. This is performed to demonstrate reweighting anthropometric data to match the population of a different country.

2.3 Different methods to obtain anthropometric data for target user population

So far, many procedures have been developed to obtain anthropometric data for the target user population. These can broadly be classified into three methods. The first is to synthesize detailed anthropometric data through estimation using the available basic anthropometric data for the target user population. Such syntheses have been performed using a variety of techniques with distinct pros and cons [30, 31, 32, 33, 34, 35]. The second method is to collect required anthropometric data from the target population. This method is common when designing for particular occupational groups [36, 37, 38]. Studies have collected anthropometric dimensions of truck drivers across the U.S. [39, 40]. These data are used in truck cabin and workspace design. In such data collections, sampling plans need to be used to capture the appropriate population characteristics such as race/ethnicity, age, and gender. The third method is to modify the available data to better represent the target user population.

The two most common approaches to modifying the available data to represent the target population are resampling and weighting. Resampling alters the number of datapoints in the data by removing or multiplying certain data to match the required overall characteristics of the target population. This approach has been used in creating the ANSUR data where data was downsampled to match the U.S. army in active duty in June 1988[11]. In weighting, each individual in the sample is assigned a weight which represents the fraction of the target population characterized by that individual. When summed, the weights equal the total number of individuals in the population. One advantage to this approach is that it makes use of all of the available data. This can improve the ability of the data to model the tails of the distribution. Weighted survey data are used in several domains such as health surveys [25] and crash statistics [41].

The National Health and Nutrition Examination Survey (NHANES) collects basic anthropometric statistics of the U.S. population [25]. Rather than gathering data by randomly sampling individuals, participants are carefully selected to ensure a good representation of the US population. Individuals in the tails of the distribution are intentionally oversampled (Figure 2.1, which improves the fidelity of the model in that region. This results in a dataset that does not directly represent the population of interest. Instead, weights must be assigned to the data so that they represents the US population.

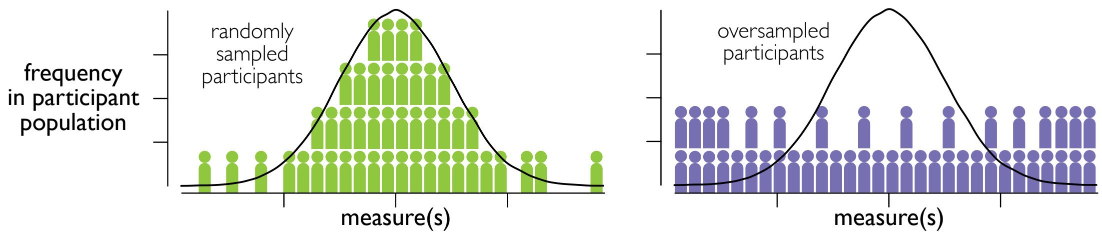

Figure 2.1: Distribution of samples obtained when different sampling techniques. The actual population distribution is shown by the curve.

2.4 Existing methods to reweight anthropometric data

Several efforts have been made in the past to assign weights to unweighted anthropometric data in order to make it representative of the target user population. The first method discussed here was used at the United States Air Force Research Laboratory where a binning approach was used to assign weights to CAESAR data using NHANES III data to represent the adult U.S. population [42]. This was performed to obtain detailed anthropometric data that represents the U.S. population of that period. Reweighting CAESAR data also provides detailed body scans that are representative of the U.S. population, which can then be used to test the accommodation of new designs. This reweighting method will be herein referred to as Harrison method in this thesis. The predictors, stature, mass, age, race, and gender were used to estimate the weights for CAESAR. A total of 45 bins each for men and women were used by binning with 2 gender groups, 3 age groups, 3 racial groups, and 5 groups based on stature and mass.

A similar attempt was made again at the United State Air Force Research Laboratory to create anthropometric data for the U.S. Air Force using CAESAR data from North America, Italy, and the Netherlands [43]. This method will be herein referred in this thesis as Hudson method. It was used to reweight CAESAR data to match the pilots of Joint Strike Fighter program of the U.S. Air Force. This provides 3D body scans which would then assist in the design and sizing of clothing and protective equipment for the pilots in this program. The CAESAR data were screened with mass regulations of Navy and Air Force, then screened with age limits and then weighted with predictions of racial composition in 2010 using three ethnicity categories: European American, African American and Asian American. Hispanic American ethnicity was excluded from the categories considered for binning as it was found that the distinct body shape and size variation which characterize the other three races was not evident in the Hispanic population[43]. Therefore it was decided that the Hispanic American population was not needed as a distinct category, as the variation was included and described by the first three mentioned ethnicities.

Between 2006 and 2008, a pilot study called ANSUR II at the United States Army Natick Soldier Research, Development and Engineering Center aimed to compare the 1988 ANSUR data with anthropometric data on the active U.S. Army in 2007. This was performed to check the effect of secular trends over the years in the U.S. Army population. Secular trends such as increase in height and BMI over the decades result in the ANSUR data being not representative of the modern U.S. army population. A binning approach was used to weight the data collected from ANSUR II to match the then demographics of U.S. army's active, reserve and National Guard components[13]. The bins were created on the basis of these components, age, race, and sex, resulting in a total of 16 bins each for men and women in each of the said three components of military duty. This method will herein be called Paquette method in this thesis.

Chapter 3

New method to reweight data

This chapter evaluates the results of reweighting anthropometric data using the existing methods. The new method is then described and the effect of different parameters on reweighting are studied. The results of reweighting with the new method are then evaluated to check the reliability of the reweighted data for design purposes. In order to compare the effectiveness of the existing weighting techniques described in the previous chapter, the North American CAESAR population is weighted using the already weighted NHANES data from 2011 to 2014. This is performed to show the requirement for a new weighting method. The errors in various statistics of the stature, mass and BMI in the reweighted CAESAR data compared to the actual NHANES data would suggest the accuracy of the reweighting method. If the errors in these three measures are large, it would signify that the percentile vales of other detailed anthropometric measures would also have similar errors. This can be inferred in cases where the anthropometric measure under consideration is proportional to stature, mass or BMI. Therefore, each of the three existing techniques are used to reweight CAESAR data, and the corresponding errors are compared.

3.1 Results of weighting CAESAR using existing weighting techniques

Table 3.1 gives the results of reweighting CAESAR data with each of the three methods described earlier. Table 3.2 gives the differences in the values obtained for original and reweighted CAESAR data from the NHANES 2011-14 data. The values compared were the 5^{th}, 50^{th}, and 95^{th} percentiles, mean, and standard deviations in three basic body measures namely, stature, mass, and BMI. The difference in stature among the different statistics was as high as 46 mm in Hudson method, 17 mm in Catherine method and 14 mm in Paquette method. Harrison method had a maximum error of 6.5 kg in mass and 1.9 in BMI among the 5 statistics considered. While Paquette method had a maximum error of 12.3 kg in mass and 4.6 in BMI among these 5 statistics. Hudson method was found to aggravate the errors in the unweighted CAESAR data upon

Table 3.1: Statistics of NHANES 2011-14 data, and North American CAESAR data reweighted using different binning methods

statistic	measure	NHANES 2011-14	unweighted CAESAR	reweighted CAESAR Harrison	reweighted CAESAR Hudson	reweighted CAESAR Paquette
5^{th} percentile	height (mm)	1534	1550	1541	1505	1535
	weight (kg)	53.9	51.8	52.2	48.1	52.6
	BMI	20.0	19.6	20.0	18.9	19.7
50^{th} percentile	height (mm)	1689	1700	1693	1645	1696
	weight (kg)	79.6	73.5	77.8	66.4	74.4
	BMI	27.6	25.2	26.3	24.2	25.4
95^{th} percentile	height (mm)	1859	1876	1876	1813	1873
	weight (kg)	123.0	113.7	116.6	102.3	110.7
	BMI	42.5	37.1	40.6	34.5	37.5
weighted mean	height (mm)	1693	1704	1700	1653	1699
	weight (kg)	83.0	76.8	80.3	69.0	77.1
	BMI	28.9	26.3	27.7	25.1	26.6
standard deviation	height (mm)	100	102	105	95	103
	weight (kg)	22.0	19.4	20.4	17.3	18.9
	BMI	7.1	5.6	6.3	4.9	5.6

reweighting. The error was maximum in the tails of the distribution for all of the methods, with higher errors below 5^{th} and above 95^{th} percentiles. Therefore, none of these methods provided a reliable way to reweight anthropometric data so as to make it suitable for use in design of products. The reweighting process had to be considerably improved before its implementation in real-world design applications.

3.2 Effect of varying number of bins in regular binning method to reweight data

All the three existing methods described in the previous section was essentially a kind of binning process. This meant that data from both the dataset representative of an actual population as well as the dataset to be reweighted were divided on the basis of one or more of its characteristics such as stature, mass and race/ethnicity. Each of the categories formed as a result of division based a single or multiple characteristic(s) is termed as a bin. Each bin would ideally contain at least one record from the representative data and at least one record from the data to be reweighted. The weights of the representative data points in each bin is summed and then divided equally amongst the records in same bin belonging to the dataset to be reweighted. This process will be termed as binning in the following sections.

One of the ways to improve the accuracy of reweighting is to increase the number of bins. This creates more number of divisions and/or divisions based on more number of characteristics.

Table 3.2: Difference in statistics of reweighted North American CAESAR data from NHANES 2011-14 data

statistic	measure	unweighted CAESAR	reweighted CAESAR Harrison	reweighted CAESAR Hudson	reweighted CAESAR Paquette
5^{th} percentile	height (mm)	16	7	29	1
	weight (kg)	2.1	1.8	5.8	1.3
	BMI	0.4	0.0	1.1	0.3
50^{th} percentile	height (mm)	11	4	44	7
	weight (kg)	6.1	1.8	13.2	5.2
	BMI	2.5	1.3	3.5	2.2
95^{th} percentile	height (mm)	17	17	46	14
	weight (kg)	9.3	6.5	20.7	12.3
	BMI	5.4	1.9	7.9	4.9
weighted mean	height (mm)	12	7	40	7
	weight (kg)	6.2	2.7	14.0	5.9
	BMI	2.6	1.2	3.8	2.3
standard deviation	height (mm)	3	5	4	3
	weight (kg)	2.6	1.6	4.7	3.1
	BMI	1.5	0.8	2.2	1.5

For example, the number of bins may be increased by reducing the range of stature in each of the bins. The number of bins may further be also increased by separating each of these bins based on the value of weight, and further based on the race, and so on. The following test was conducted to evaluate the improvement in reweighting process upon increasing the number of bins.

In this study, bins were created based on two characteristics namely, stature and mass. The number of bins on the two characteristics were both increased simultaneously from 2 to 40 in steps of one. In each step, the data was reweighted using the earlier described method of binning. Therefore, the statistical weights of records from the representative data were redistributed among those of the dataset to be reweighted, within the same bin. Increasing the number of bins in each of the two characteristics simultaneously, resulted in an overall increase in the number of bins from 4 to 400 in 38 steps. This followed the pattern $2^2, 3^2, 4^2, ...40^2$. In each of these stages, each of the created bins were checked whether they contained at least one data point from each of the two datasets. In cases where it did not satisfy this condition, it was merged with the nearest bin which when combined with, contained at least one data point from each of the datasets, in the resulting merged bin.This ensured that all the sum of statistical weight in the representative data and the reweighted data set amounts to the same value. In other words, the number of people represented by both the datasets would be equal, upon reweighting. This process reduced also the number of bins from the initial count in most cases, resulting in a maximum of 849 bins formed in this study.

Once the reweighting process was performed, the percentiles values of 1 to 100 were calculated

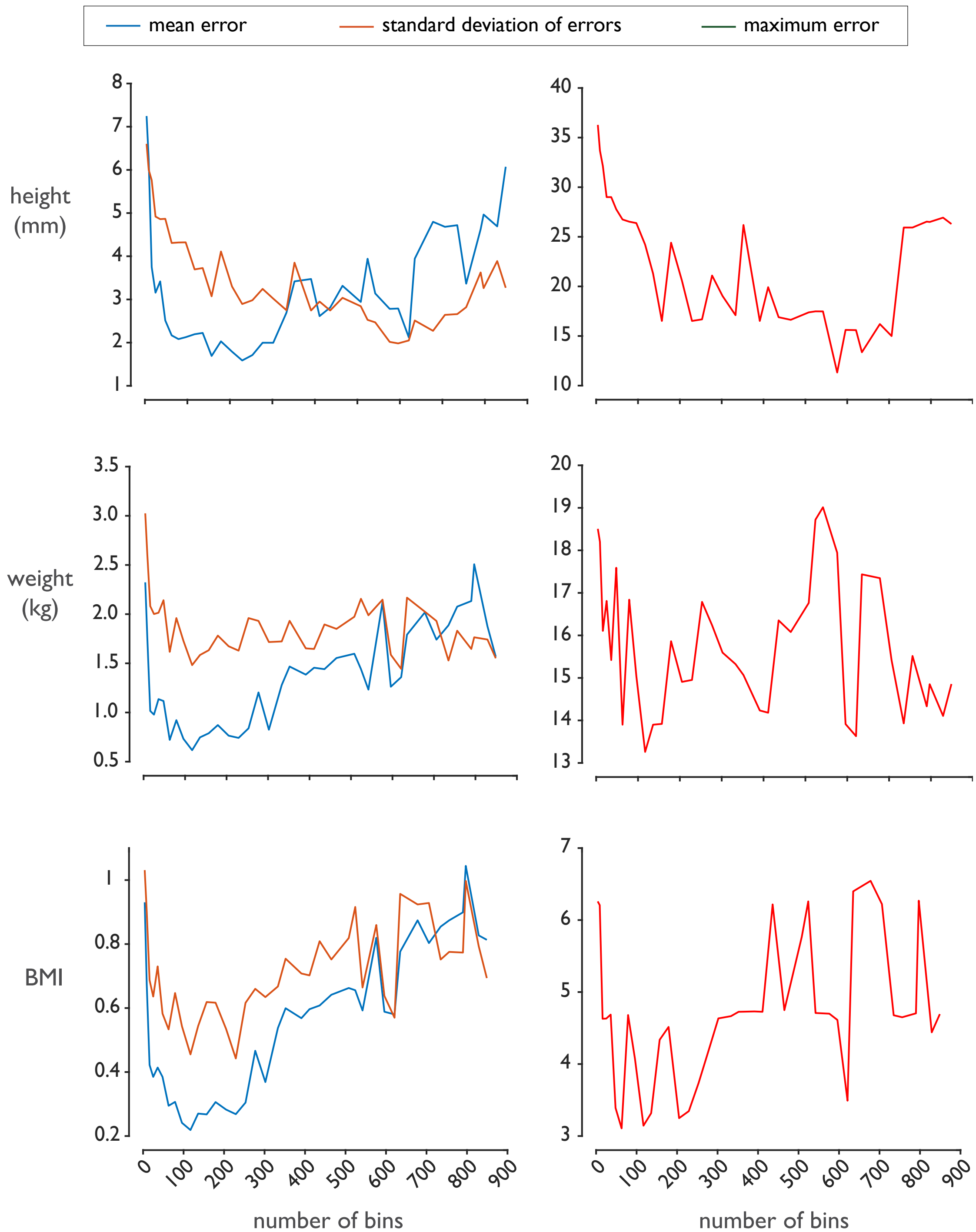

Figure 3.1: Mean, standard deviation and maximum errors in percentiles 1-100 of Stature, Mass and BMI, on comparing reweighted CAESAR Male data with NHANES 11-14 Male data.

for each of the three body measures, stature, mass and BMI. The differences of these percentile values from the corresponding percentiles of the representative dataset were termed as errors.

These errors were calculated for male CAESAR data reweighted with each of the 38 binning strategies. A total of 100 error values at each percentile value (1 to 100) were calculated for each of the three body measures considered. Figure 3.1 shows the variation of mean, standard deviation and maximum of these errors with the number of bins used in the reweighting process. It was found that upon increasing the number of bins from 2 to around 10 along each of the two characteristics, stature and mass, the errors had a decreasing trend, omitting certain outliers. But upon further increasing the number of bins, the errors also increased. Therefore, it was observed that mere increase in the number of bins did not improve the accuracy of reweighting beyond a particular level. Further improvement in the reweighting process is required before reweighted data can be used in design of products.

3.3 New method to reweighting anthropometric data

In the previous section, it was observed that the errors with the reweighted data were large with all the existing methods that were compared. Therefore, a new method is required which can reweight anthropometric data more reliably. The following section describes this new method that is introduced in this thesis. It is later evaluated to find the errors involved in the reweighted data. In cases where the data to be weighted already contained weights, these weights were removed. The data to be weighted or reweighted is herein called unweighted data. The word reweighted is used in this thesis regardless of whether the data to be reweighted already contained weights or not. The data representative of a the target population may or may not contain weights. In cases where this representative data did not contain weights, each of its data points were assigned the same weight of 1, and therein called weighted data. In all cases, the data for males and females were weighted separately and then combined as needed. This can be considered the same as having gender as an additional binning criteria. This was necessary due to the anthropometric differences between males and females.

3.3.1 Binning

The data were first sorted on gender. In this reweighting method, each bin should contain a minimum of one record from the weighted data and one record from the data to be weighted. Each bin had to contain both of these records as the weight(s) of the weighted record(s) were used to assign weight(s) to the remaining record(s) in the same bin. For this purpose, each record in the unweighted data was assigned a unique bin number, resulting in the number of bins being equal to the number of records in the unweighted data. Therefore,

$$b = N_0, \tag{3.1}$$

where b is the bin number and N_0 is the record number in the unweighted data. Every record in the weighted data needs a bin associated with it where its weights can be redistributed. So, for each of the records in the weighted data, the nearest unweighted record was found by calculating the Euclidean distance between them. The bin number of this unweighted record was then assigned to the weighted data record. The Euclidean distance, D between each pair of records was calculated as,

$$D = \sqrt{\left(\frac{S_1 - S_2}{R_S}\right)^2 + \left(\frac{M_1 - M_2}{R_M}\right)^2 + \left(\frac{B_1 - B_2}{R_B}\right)^2}, \tag{3.2}$$

where S denotes stature, M denotes mass, B denotes BMI, and R_S, R_M, and R_B denotes the ranges of stature, mass, and BMI respectively of either of the datasets. The range of stature, mass, and BMI were used in order to find the normalized distance using these three characteristics of the two data points whose distance from each other was being calculated. For this purpose both the ranges of variables in the weighted data and the unweighted data were used in separate cases to identify its effect on the accuracy of the final weights. The bins which still did not contain at least one record from the weighted data, were to be merged with other bins. These records were instead given the bin number of the nearest record from the weighted data. Equation 2 was

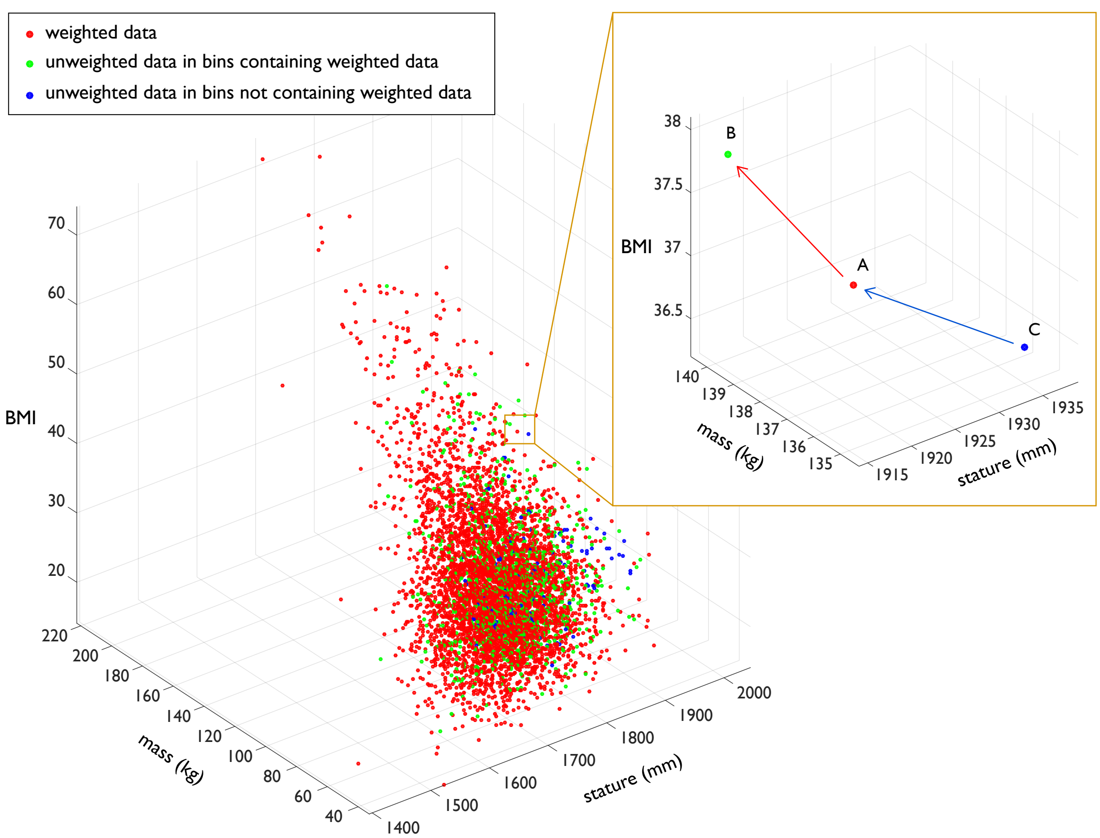

Figure 3.2: An Example of weighted and unweighted data are illustrated in a 3D graph of their stature, mass and BMI. The zoomed in region shows the weighted data point A assigned the bin of it closest unweighted data point B, and unweighted data point C assigned the bin of its closest weighted data point A since its bin does not contain any weighted point from the first round of bin assignments.

again used for this purpose. 3.2 shows a scenario where the weighted data point A is assigned to the bin of point B in the unweighted data since B is the unweighted data point closest to A. The bin associated with Point C in the unweighted data does not contain a weighted data point and is therefore assigned to the bin associated with point A which is the closest weighted data point.

3.3.2 Assignment of weights

After the binning process was completed in this manner, the sum of weights of weighted records in each of the bins were equally divided among the remaining records in the same bin according to the equation,

$$W_0 = \frac{\sum W_w}{n_0}, \tag{3.3}$$

where $\mathrm{W_w}$ denotes the weights assigned to the weighted records in the bin, $\mathrm{W_0}$ denotes the weight to be assigned to each of the remaining records in the same bin, and $\mathrm{n_0}$ denotes the number of records from the unweighted data which are assigned the same bin number as that which was being considered.

3.4 Variations of the new weighting technique

The new method introduced in the previous section can be implemented in a variety of ways. A different set of measures can be used to find matching pairs between the weighted and unweighted data in place of using all the three basic measures, stature, mass, and BMI. Different ranges may be used for normalizing the stature, mass, and BMI, thereby changing the bins formed in the reweighting process. This section looks at the effects of such variations in order to minimize the error observed in the reweighted data (unweighted data after weighting), when compared to the actual weighted data.

3.4.1 Effect of using BMI in finding euclidean distance

The new reweighting technique introduced here uses three body measures namely, stature, mass and BMI to find the euclidean distance between data points of representative dataset and those of the data to be weighted. BMI is derived from stature and mass. The following steps prove that the inclusion of BMI will alter the assignment of bins to the data points, although BMI is a function of the other two body measures considered. Equation 3.2 changes to the following equation 3.4 when BMI is not included in the calculation of distance.

$$D = \sqrt{\left(\frac{S_1 - S_2}{R_S}\right)^2 + \left(\frac{M_1 - M_2}{R_M}\right)^2}, \tag{3.4}$$

where S denotes stature, M denotes mass, and $\mathrm{R_S}$ and $\mathrm{R_M}$ denote the ranges of stature and mass respectively of either of the datasets. If the inclusion of BMI has to alter the bin assignment, the closest data point to a reference data point has to be different in cases where BMI is considered, and not considered in equation 3.2 for finding the euclidean distance. It is observed that for a set of three points, $A(x_1, y_1, z_1)$, $B(x_2, y_2, z_2)$ and $C(x_3, y_3, z_3)$, where x, y and z denote stature, mass and BMI respectively, it is possible that the distance between C and A may be less than the distance between B and A when BMI is not considered (when using equation 3.4), but is greater than the distance between B and A when BMI is considered in the equation to find euclidean distance (when using 3.2). This is the case when the three points A, B, and C satisfies the two conditions, given in equation 3.8 and 3.5.

$$(z_1 - z_2)^2 + (z_1 - z_3)^2 > (x_1 - x_3)^2 + (y_1 - y_3)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2 \tag{3.5}$$

When z denotes BMI, z changes to,

$$z = \frac{y}{x^2} \tag{3.6}$$

since,

$$\text{BMI} = \frac{\text{Mass in kilograms}}{(\text{Stature in meters})^2} \tag{3.7}$$

Given three points,

$$A\left(x_1, y_1, \frac{y_1}{(x_1)^2}\right)$$
$$B\left(x_2, y_2, \frac{y_2}{(x_2)^2}\right)$$
$$C\left(x_3, y_3, \frac{y_3}{(x_3)^2}\right)$$

Point C is closer to point A than point B when BMI is considered along with stature and mass in finding the euclidean distance, but not when BMI is not considered, if the three points satisfy the following two conditions (equations 3.8 and 3.9).

Condition 1:

$$(x_1 - x_2)^2 + (y_1 - y2)^2 < (x_1 - x_3)^2 + (y_1 - y_3)^2 \tag{3.8}$$

Condition 2:

$$\left(\frac{y_1}{(x_1)^2} - \frac{y_2}{(x_2)^2}\right)^2 + \left(\frac{y_1}{(x_1)^2} - \frac{y_3}{(x_3)^2}\right)^2 > (x_1 - x_3)^2 + (y_1 - y_3)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2 \tag{3.9}$$

Proof.

Condition 1:

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 < (x_1 - x_3)^2 + (y_1 - y_3)^2$$

Let 2D distance represent the Euclidean distance calculated using only stature and mass, while

3D distance represent the Euclidean distance calculated using stature, mass and BMI.

$$\Big(\text{2D distance between point A and B}\Big)^2 < \Big(\text{2D distance between points A and C}\Big)^2$$

Distance is never negative. So,

$$\text{2D distance between point A and B} < \text{2D distance between points A and C}$$

Condition 2:

$$\left(\frac{y_1}{(x_1)^2} - \frac{y_2}{(x_2)^2}\right)^2 + \left(\frac{y_1}{(x_1)^2} - \frac{y_3}{(x_3)^2}\right)^2 > (x_1 - x_3)^2 + (y_1 - y_3)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2$$

Using equation 3.6, we get,

$$(z_1 - z_2)^2 + (z_1 - z_3)^2 > (x_1 - x_3)^2 + (y_1 - y_3)^2 - (x_1 - x_2)^2 - (y_1 - y_2)^2$$

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 > (x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2$$

$$\Big(\text{3D distance between point A and B}\Big)^2 > \Big(\text{3D distance between points A and C}\Big)^2$$

Distance is never negative. So,

$$\text{3D distance between point A and B} > \text{3D distance between points A and C}$$

Therefore, for three points A, B, and C, it can be shown that point B is closer than point C to point A when x and y (stature and mass) alone are considered in the distance equation, whereas point C is closer than point B to point A when z (BMI) is also considered in the distance equation, by proving that these points satisfy assumptions 1 and 2. Consider the example where there are three points,

$$A(X_1, Y_1, Z_1) = (1.65, 83.0, 30.49)$$
$$B(X_2, Y_2, Z_2) = (1.60, 82.4, 32.19)$$
$$C(X_3, Y_3, Z_3) = (1.70, 84.0, 20.07)$$

X, Y, and Z represent stature in meters, mass in kilograms, and BMI respectively. Here Z being BMI, satisfies equation 3.6.

Substituting these in condition 1,

$$(1.65 - 1.6)^2 + (83 - 82.4)^2 < (1.65 - 1.7)^2 + (83 - 84)^2$$

$$0.36 < 1.00$$

Condition 1 is true.

Substituting these in condition 2,

$$(30.49 - 32.19)^2 + (30.49 - 29.07)^2 > (1.65 - 1.7)^2 + (83 - 84)^2 - (1.65 - 1.6)^2 - (83 - 82.4)^2$$

$$4.91 > 0.64$$

Condition 2 is true.

This example set of points prove that both assumptions 1 and 2 can be true for the same set of points. The following steps show that this cannot be avoided by normalizing the values of X, Y and Z with their respective ranges in the dataset. Since X, Y and Z represent stature, mass and BMI, the ranges of these were taken from an anthropometric database, NHANES 2011-14 adult males.

$$\text{Range of stature} = 0.65 \text{ m}$$
$$\text{Range of mass} = 189.8 \text{ kg}$$
$$\text{Range of BMI} = 60.0$$

The following normalized values x, y, and z were obtained using the ranges in NHANES 2011-14 male data.

$$\begin{aligned} A_n(x_1, y_1, z_1) &= \left(\frac{1.65}{0.65}, \frac{83}{189.8}, \frac{30.49}{60}\right) \\ &= (2.54, 0.44, 0.51) \\ B_n(x_2, y_2, z_2) &= \left(\frac{1.60}{0.65}, \frac{82.4}{189.8}, \frac{32.19}{60}\right) \\ &= (2.46, 0.43, 0.54) \\ C_n(x_3, y_3, z_3) &= \left(\frac{1.70}{0.65}, \frac{84.0}{189.8}, \frac{20.07}{60}\right) \\ &= (2.62, 0.44, 0.49) \end{aligned}$$

Substituting these in condition 1,

$$\left(\frac{1.65}{0.65} - \frac{1.60}{0.65}\right)^2 + \left(\frac{83}{189.8} - \frac{82.4}{189.8}\right)^2 < \left(\frac{1.65}{0.65} - \frac{1.70}{0.65}\right)^2 + \left(\frac{83}{189.8} - \frac{84.0}{189.8}\right)^2$$

$$0.00593 < 0.00594$$

Condition 1 is true.

Substituting these in condition 2,

$$\begin{aligned} \left(\frac{30.49}{60} - \frac{32.19}{60}\right)^2 + \left(\frac{30.49}{60} - \frac{20.07}{60}\right)^2 > &\left(\frac{1.65}{0.65} - \frac{1.70}{0.65}\right)^2 + \left(\frac{83}{189.8} - \frac{84.0}{189.8}\right)^2 \\ &- \left(\frac{1.65}{0.65} - \frac{1.60}{0.65}\right)^2 - \left(\frac{83}{189.8} - \frac{82.4}{189.8}\right)^2 \end{aligned}$$

$$0.00024 > 0.00002$$

Condition 2 is true.

Now in order to prove that this is possible even when the values of stature, mass and BMI are normalized to vary between 0 and 1, these values were normalized by dividing using their respective maximum values obtained from NHANES 2011-14 adult male data. We consider a different set of three points,

$$A(X_1, Y_1, Z_1) = (1.833, 103.90, 30.9236)$$
$$B(X_2, Y_2, Z_2) = (1.838, 103.40, 30.6076)$$
$$C(X_3, Y_3, Z_3) = (1.839, 104.76, 30.9765)$$

These maximum values used for normalization were,

$$\text{Maximum stature} = 2.045 \text{ m}$$
$$\text{Maximum mass} = 222.6 \text{ kg}$$
$$\text{Maximum BMI} = 74.1$$

After normalizing the values of points A, B, and C, the following points were obtained.

$$\begin{aligned}
A_n(x_1, y_1, z_1) &= \left(\frac{1.833}{2.045}, \frac{103.9}{222.6}, \frac{30.9236}{74.1}\right) \\
&= (0.8963, 0.4668, 0.4173) \\
B_n(x_2, y_2, z_2) &= \left(\frac{1.838}{2.045}, \frac{103.4}{222.6}, \frac{30.6076}{74.1}\right) \\
&= (0.8988, 0.4645, 0.4131) \\
C_n(x_3, y_3, z_3) &= \left(\frac{1.839}{2.045}, \frac{104.76}{222.6}, \frac{30.9765}{74.1}\right) \\
&= (0.8993, 0.4706, 0.4180)
\end{aligned}$$

Substituting these in condition 1,

$$(1.833 - 1.838)^2 + (103.90 - 103.40)^2 < (1.833 - 1.839)^2 + (103.90 - 104.76)^2$$
$$0.0001 < 0.00002$$

Condition 1 is true.

Substituting these in condition 2,

$$\begin{aligned}
(30.9236 - 30.6076)^2 + (30.9236 - 30.9765)^2 > (1.833 - 1.839)^2 + (103.90 - 104.76)^2 \\
- (1.833 - 1.838)^2 - (103.90 - 103.40)^2
\end{aligned}$$
$$0.000018 > 0.000013$$

Condition 2 is true.

It is seen that there exists combinations of stature, mass and BMI which satisfies the two conditions. That is, there exists combinations of points where a point C is closer to A than B when BMI is considered in the equation for euclidean distance, but farther than B when BMI is not considered. Therefore it is proved that the closest point for a reference point can be different when BMI is considered and when BMI is not considered in the Euclidean distance equation, even with normalization, although BMI is derived from stature and mass. So the created bins are also different in the two cases. This is the reason why BMI was not omitted in equation 3.2 for calculating euclidean distances. Body measures other than stature, mass and BMI were still not included since they are not available in most anthropometric data. □

3.4.2 Effect of different ranges used for normalization during reweighting

For the purpose of evaluation of the proposed weighting technique, CAESAR data on men from North America were weighted using the NHANES 2011-14 data on men aged from 18 to 65. This age group was chosen since CAESAR data contained anthropometric measures of people in this age group. The NHANES data used, represent the male population of the US from year 2011 to 2014. Whereas, the male CAESAR data used was a very detailed anthropometric database which did not represent an actual population. Several characteristics of stature, mass, and BMI in the reweighted CAESAR data were analyzed in order to evaluate how representative these reweighted data were of the 2011-14 US population. The weighted mean and standard deviation of each of the variables provided an overall estimate of the population. In addition to these, the 5$^{\text{th}}$, 50$^{\text{th}}$, and 95$^{\text{th}}$ percentile values were also compared as these values are crucial to several design decisions. Table 3.3 shows the comparison of these values between CAESAR data reweighted

Table 3.3: Statistics of newly weighted CAESAR men datasets and NHANES 2011-14 men dataset

statistic	measure	reweighted CAESAR method 1	reweighted CAESAR method 2	reweighted CAESAR method 3	weighted NHANES 2011-2014
5$^{\text{th}}$ percentile	stature (mm)	1638	1638	1638	1639
	mass (kg)	61.9	61.9	61.9	61.7
	BMI	20.6	20.6	20.6	20.6
50$^{\text{th}}$ percentile	stature (mm)	1762	1761	1761	1761
	mass (kg)	85.9	85.9	85.9	86.0
	BMI	27.6	27.6	27.6	27.7
95$^{\text{th}}$ percentile	stature (mm)	1883	1883	1885	1885
	mass (kg)	127.4	128.1	128.1	127.8
	BMI	40.4	40.4	40.4	40.1
weighted mean	stature (mm)	1763	1763	1763	1762
	mass (kg)	88.9	88.9	88.9	89.0
	BMI	28.6	28.6	28.6	28.6
standard deviation	stature (mm)	74.4	74.2	74.2	75.8
	mass (kg)	20.5	20.4	20.4	20.9
	BMI	6.0	6.0	6.0	6.2

Table 3.4: Mean and maximum absolute errors among percentile values of reweighted CAESAR men datasets

statistic	measure	reweighted CAESAR method 1	reweighted CAESAR method 2	reweighted CAESAR method 3
mean absolute error	stature (mm)	1	1	1
	mass (kg)	0.3	0.3	0.3
	BMI	0.1	0.1	0.1
maximum absolute error	stature (mm)	14	16	15
	mass (kg)	7.0	7.7	7.3
	BMI	1.8	1.4	1.7

using three different sets of ranges in equation 3.2, and the original weighted NHANES 2011-14 men data. Method 1 uses the stature, mass and BMI ranges of the weighted NHANES data. Method 2 uses those ranges of the unweighted CAESAR data. Method 3 uses the ranges between 10^{th} and 90^{th} percentiles of stature, mass, and BMI of the weighted NHANES data. Method 3 is considered to avoid the extreme values in the three measures. None of these ranges may provide the optimum value for normalizing the values of stature, mass, and BMI for finding the unitless euclidean distance, but the results can still be compared to find the better option. It is seen that the values of different statistics of the three body measures are very similar between the three methods in table 3.3.

In order to better evaluate the weighting, percentiles 1 to 99 of stature, mass, and BMI were calculated for the reweighted CAESAR data and compared with those of the NHANES 2011-14 data. The difference in the values of each of these percentiles between the weighted CAESAR data and NHANES data were termed as errors. The maximum and mean of these absolute errors were compared for the weighting technique using the three different ranges for normalization mentioned earlier. Table 3.4 shows this comparison of mean and maximum absolute error in the percentiles. It is to be noted that the maximum absolute error occurs mostly only at the highest or lowest percentile of the data. Here too its seen that the range chosen for the reweighting method did not affect the errors considerably. Figure 3.3 shows the absolute errors i.e. the absolute difference between weighted CAESAR data and NHANES data at percentiles 1 to 99 of each of the variables, stature, mass, and BMI.

The choice of using the ranges of the variables in the weighted (NHANES) or unweighted (CAESAR) datasets to normalize the distance between data points, as in equation 3.2 was not found to have a noticeable impact on the statistical measures of the final data. It can be seen in Table 3.4 that the mean absolute error among the 99 percentiles were almost the same whereas the maximum absolute errors were not clearly better on either of the normalization methods. No noticeable difference in the errors are seen visually (refer Figure 3.3). The statistical measures such as weighted mean and standard deviation of the weighted CAESAR data were found to be very close to those of NHANES 2011-14 data (refer Table 3.3).

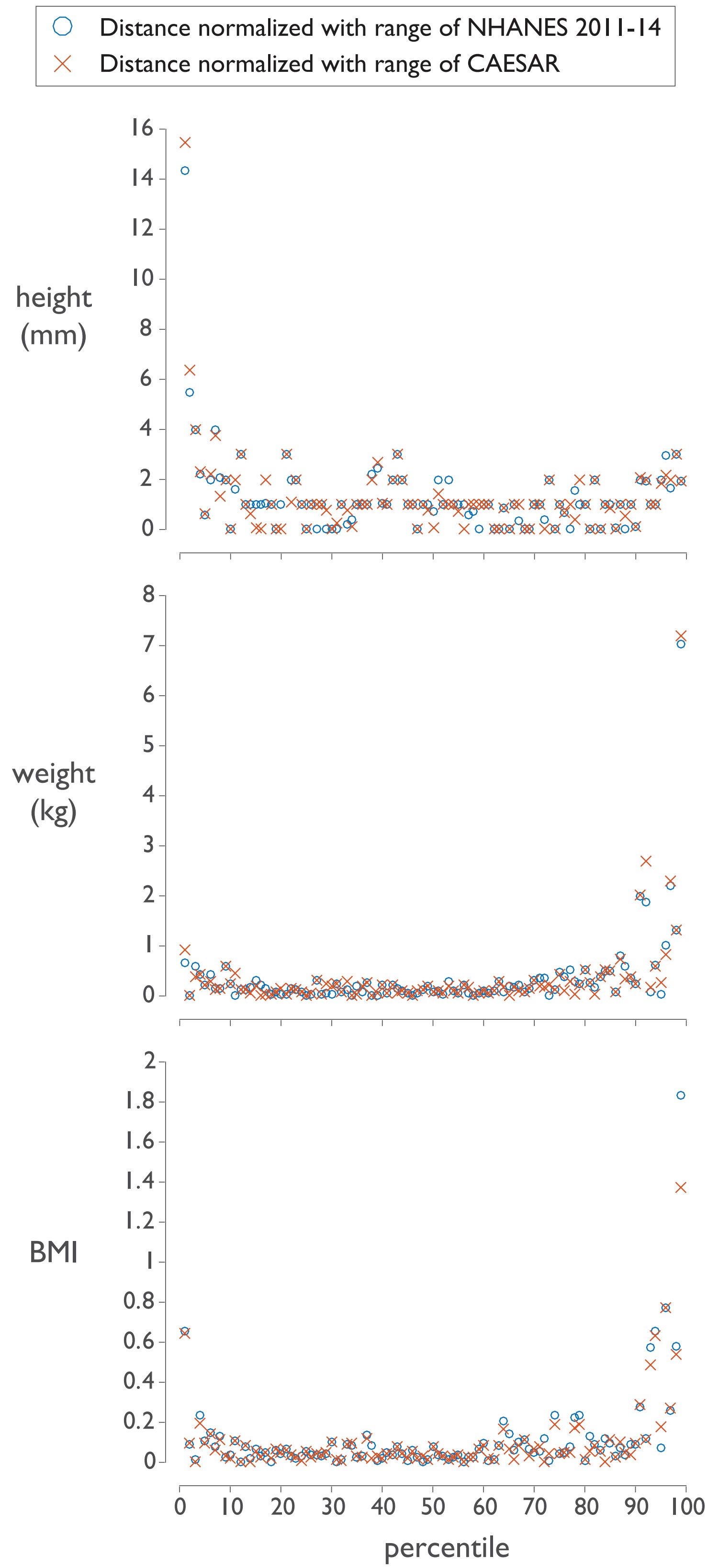

Figure 3.3: Absolute errors at each percentile of newly weighted CAESAR data

3.4.3 Further evaluation of reweighted data

In the following comparisons, the weighting were all performed using Method 1 where the ranges from the weighted dataset were used in equation 3.2. Nine hundred and ninety three bins were formed when weighting CAESAR data using NHANES data. The Figure 3.5 shows the values associated with percentiles 1 to 99 of stature, mass, and BMI of the weighted CAESAR data (using method 1) superimposed on those of NHANES 2011-14 data. The vertical distance between the two at each percentile shows the error in the reweighted data at that percentile. It is seen that these point almost coincide at each percentile values with any errors visible only at the tails of the distribution. The percentile values of mass and BMI were visually seen to be less accurate at only the three to five percentiles in the higher tail. The mean and maximum absolute errors among the percentiles are given in table 3.4.

Apart from the statistical measures, weighted covariance matrices were used to compare the relationships between variables in weighted CAESAR data and the relationships between variables in NHANES data. In order to eliminate the effect of reduced sample size in CAESAR compared to NHANES, a random subset of NHANES with a sample size same as that of CAESAR was created with their weights so that the overall distribution is similar to the original NHANES data. The covariance matrix of this downsampled NHANES data was calculated. This procedure was performed 10000 times for a stochastic analysis. The mean and standard deviation of these 10000 matrices were then calculated. The mean of the covariance matrices obtained from the downsampled NHANES and the covariance matrix of the weighted CAESAR data were compared to calculate the number of standard deviations by which they differed (refer table 3.5). It is observed that the number of standard deviations do not exceed 1 for any of the relations. This is illustrated in figure 3.4 where the kernel density plot of each element of the covariance matrix of downsampled NHANES data is shown, with the corresponding value from the covariance matrix of reweighted CAESAR data superimposed on it.

Table 3.5: Comparing weighted covariance matrices

(a) Mean: NHANES

	stature	mass	BMI
stature	57.32	58.85	0.49
mass	58.85	436.58	120.42
BMI	0.49	120.42	38.54

(b) Standard deviation: NHANES

	stature	mass	BMI
stature	2.86	5.94	1.66
mass	5.94	30.50	8.71
BMI	1.66	8.71	2.73

(c) Weighted CAESAR

	stature	mass	BMI
stature	55.41	59.42	1.50
mass	59.42	418.20	114.12
BMI	1.50	114.12	35.98

(d) Number of standard deviations

	stature	mass	BMI
stature	0.67	0.10	0.61
mass	0.10	0.60	0.72
BMI	0.61	0.72	0.94

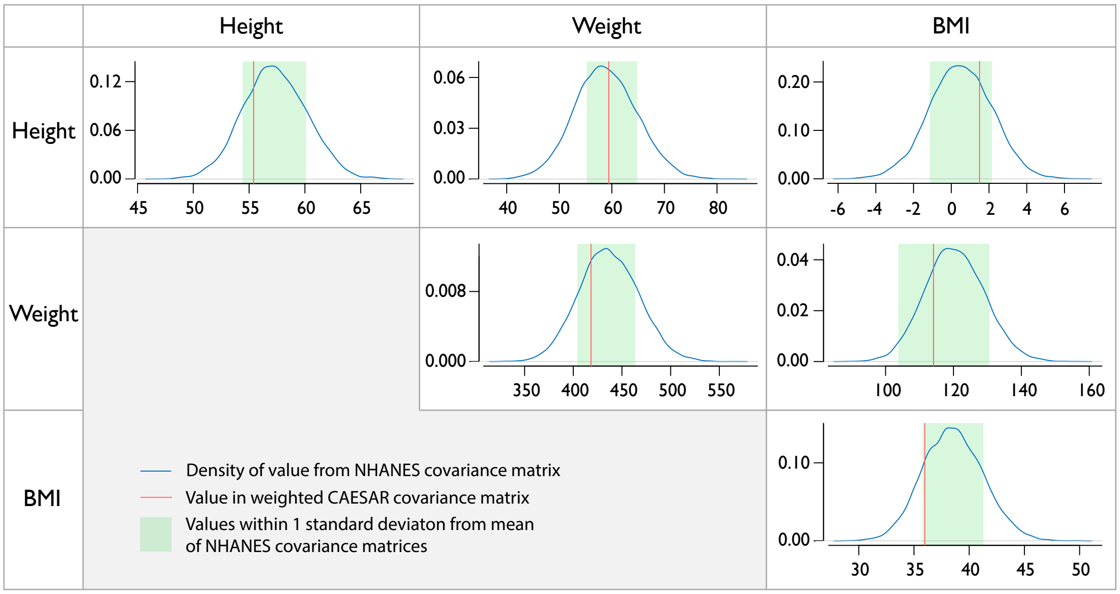

Figure 3.4: Kernel density plot of each cell in the covariance matrix of downsampled NHANES data; X axis is covariance, Y axis is kernel density of each covariance value from the stochastic study of downsampled NHANES data

3.4.4 Effect of additional binning criteria and sample size of unweighted data

In order to evaluate the effect of the sample size of the unweighted data on the performance of the weighting technique, the size of the sample to be weighted was varied between 100 and 2000 in progressions of 100. Due to unavailability of sample size of 2000 from the CAESAR male data, NHANES 2007-2010 adult male data were used after removing its statistical weights. Therefore, random samples of varying sizes picked from unweighted NHANES 2007-10 adult male data were weighted using the weighted NHANES 2011-14 adult male data. The percentiles 1 to 99 of stature, mass, and BMI of the reweighted NHANES 2007-10 data were compared with those of NHANES 2011-14 data. The difference of these percentiles were termed as errors. The maximum, mean, and standard deviation of these absolute errors were compared for different sample sizes of the unweighted data to analyze the effect of sample size on weighting. Figure 3.6 shows the maximum, mean, and standard deviation of errors at percentiles 1 to 99 of the three variables in NHANES 2007-10 data weighted using three different methods. The first method was the basic weighting method introduced in this paper, as described in the section 3.3 (using ranges of variables in NHANES 2011-14). The second method reweighted data similar to the first method, but it used an additional binning criteria of race. Three racial groups were used for binning, namely non-Hispanic White, non-Hispanic Black, and other races. The third method used was the Harrison method described in section 2.4 [42].

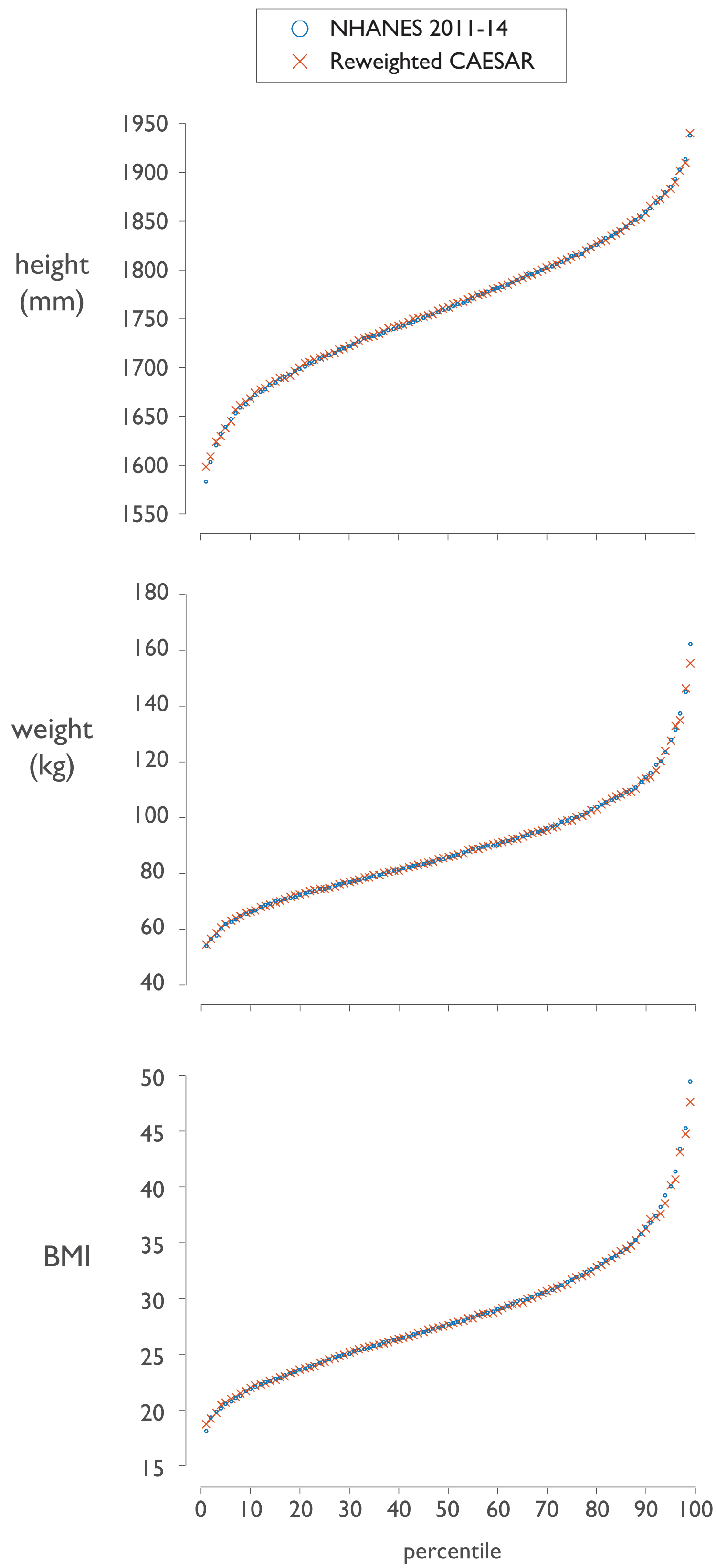

Figure 3.5: Percentiles of three body measures in NHANES 2011-14 and newly weighted CAESAR datasets

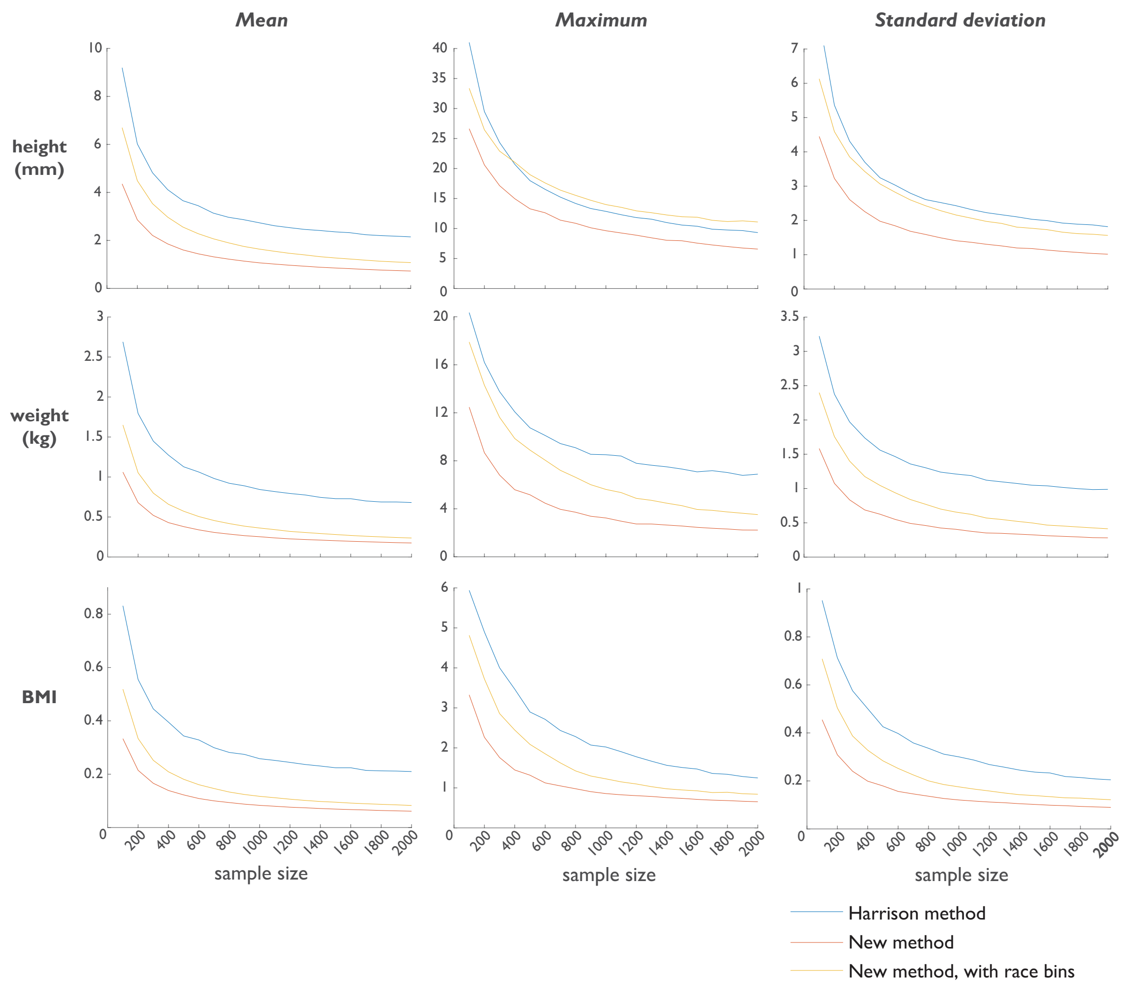

Figure 3.6: Statistics of error in reweighted data when rewighted using different sample sizes of actual weighted data

3.4.4.1 Statistical study to identify whether the reduction in error with the new methods were statistically significant

Mann-Whitney U test was performed to find the statistical significance of the difference in errors between three reweighting methods illustrated in Figure 3.6. The statistical significance was tested for sample size of 2000. This was because, as seen in Figure 3.6, the difference between the mean of the errors at each percentile value, are seen to lower with increasing sample size. The sample size was not increased beyond 2000, since this sample was downsampled from the actual NHANES 2011-14 adult male data which contained only 4471 observations. As the sample size increases, the variation in the randomly downsampled population decreases. Therefore, when the entire 4471 observations are used, the same error values will be obtained in all iterations of reweighting.

In order to conduct Mann-Whitney U test to check the statistical significance of the difference in mean error in BMI for each method, the weighted NHANES 2011-14 male adult data was randomly downsampled to 2000 observations and used to reweight an unweighted NHANES 2007-10 male adult data. This was repeated 10000 times as a Monte Carlo simulation. For each iteration, the errors in height, weight and BMI were recorded at each percentile value from 1 to 99. The mean of these errors were recorded for each of the 10000 iterations. These 10000 mean errors of BMI obtained when using each of the three methods, the new method introduced in section 3.3, the same method with an additional binning criteria of race, and the Harrison method described in section 2.4, were tested for statistically significant difference. This was conducted only for BMI, since if one of the three measures, stature, mass and BMI are found to be significantly different for the methods, then that gives reason to prefer one method over the other.

All statistical tests were conducted in the software, IBM SPSS Statistics, version 24. The dataset obtained from the three methods were tested for normality using Kolmogorov-Smirnov test and the results are shown in Table 3.6. Here the null hypothesis,

H_0: The sample data are not significantly different than a normal population.

The alternate hypothesis,

H_1: The sample data are significantly different than a normal population.

For all the three methods or conditions, the Kolmogorov-Smirnov test gave a p-value less than 0.05 with a degree of freedom (dof) of 10000. Therefore the null hypothesis can be rejected for all three conditions, showing that none of the datasets were normally distributed.

Table 3.6: Results of Kolmogorov-Smirnov test for normality on the mean errors in BMI, for NHANES 2007-10 male adult data reweighted using three different methods.

method used to reweight	statistic	degrees of freedom	significance
Harrison	0.030	10000	< 0.0005
New method	0.010	10000	0.015
New, with race bins	0.009	10000	0.047

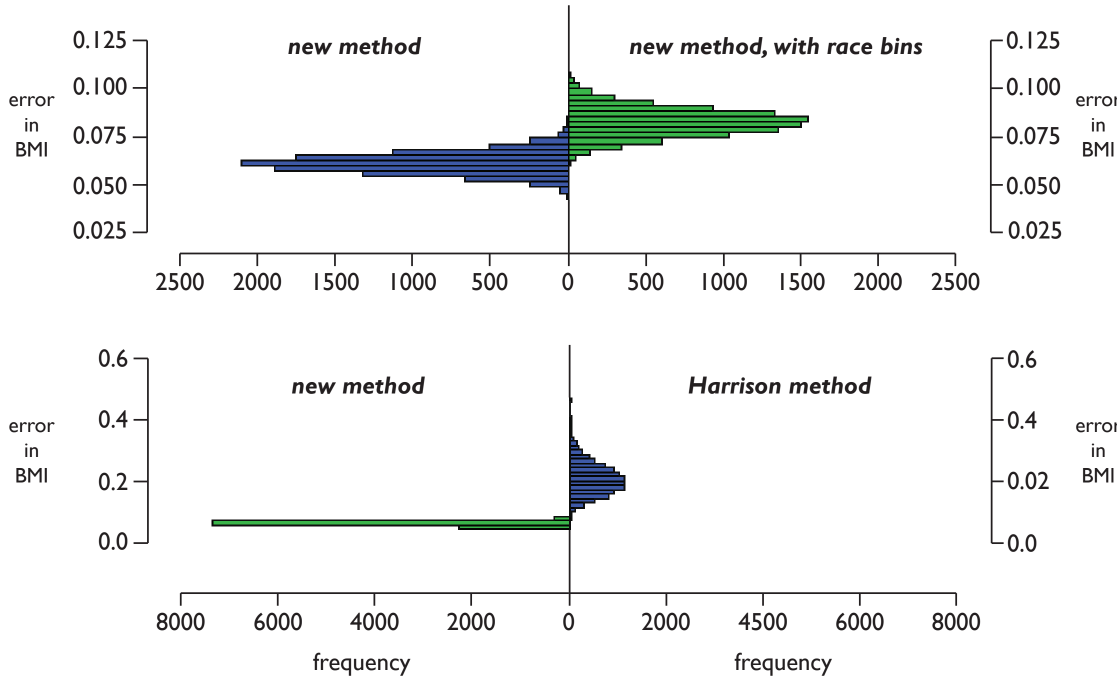

Figure 3.7: Population pyramids to compare the distribution shape of the errors in BMI obtained with the new method, to those obtained with the new method with race bins, and the Harrison method.

Using this test requires the study design and the data to fulfil certain assumptions. The data on mean errors in BMI obtained from the stochastic study contained a continuous dependent variable namely the mean error in each iteration. The independent variable was the method used for reweighting, which has three groups out of which only two groups were considered at a time for the statistical study. The observed values of errors are independent of each other. These three statements satisfy the required assumptions about the study design required for the Mann-Whitney U test.

The data distribution was then compared to check if they have similar distributions. Figure 3.7 shows the populations pyramids of the the errors in BMI obtained with the three methods. Therefore, the errors associated with these two methods can be compared using their medians. A Mann-Whitney U test was run to determine if there were differences in mean error among the percentiles of BMI in the reweighted data, when using the new method, and the new method with race bins. Distributions of the error values for these two methods were similar, as assessed by visual inspection (refer figure 3.7). Median value of mean errors associated with the new method with race bins (0.083) was higher than that obtained with the new method without race bins (0.061), $U = 145$, $z = -3.422$, $p < 0.0005$.

A Mann-Whitney U test was run to determine if there were differences in mean error among the percentiles of BMI in the reweighted data, when using the new method and Harrison method. Distributions of the error values for these two methods were not similar, as assessed by visual inspection (refer figure 3.7). Error values associated with Harrison method (mean rank = 15,000.5) were statistically significantly higher than those associated with the new method (mean rank = 5,000.5), $U = 218$, $z = -3.422$, $p < 0.0005$.

3.4.4.2 Inferring the effect of sample size of actual weighted data and additional binning criteria

It was shown that the mean errors in percentiles of BMI was statistically significantly different when using the new method compared to both the new method with race bins, and Harrison method. Several anthropometric measures are correlated with the BMI. So those anthropometric measures would also be significantly different between the methods, thereby not necessitating the need for additional statistical studies on errors in stature or mass. It was therefore observed that adding an additional constraint increased the error in re-weighting. Since the statistical study was conducted at sample size of 2000, where the difference between the methods were minimal (as seen in figure 3.6), it is safe to assume that the three methods gave significantly different results even at lower sample sizes. The maximum, mean, and standard deviations of the errors on comparing the re-weighted NHANES 2007-10 data with the NHANES 2011-14 data were found to decrease with increase in the unweighted sample size as observed by visual inspection of Figure 3.6). This shows that higher the sample size, higher the accuracy of weighting. The effect of increase in sample size was found to decrease with the increase in the sample size. On comparing results obtained with those obtained using the Harrison method, it was found that the maximum, mean, and standard deviation of the errors in stature, mass, and BMI were much lower with the new technique. The reduction in error was more pronounced in mass and BMI in this particular case of reweighting NHANES data.

Chapter 4

Reweighting different datasets using the new method

One of the most important applications of reweighting is to obtain reliable anthropometric data for countries where detailed anthropometric surveys have not been performed to obtain data representative of the population. Similar to obtaining data for different countries, it is equally desirable to obtain anthropometric data for a specialized population. Army, Air Force pilots, and truck drivers are some of such specialized populations which have been subjects of research to better design products for. This section looks at reweighting available suitable detailed anthropometric data for similar desirable populations. For the comparison of reweighted and existing representative samples, 5 statistical measures were compared. The 5^{th}, 50^{th}, and 95^{th} percentile values were considered as these values influence important design decisions in most cases. The mean and standard deviations were compared as these give an overall measure of similarity between the two datasets. Each of these statistics were compared for stature, mass and BMI, as many anthropometric measures are proportional to one or more of these three measures.

Table 4.1 shows different statistical measures of NHANES 2011-14 adults (ages 18-65) data and CAESAR data reweighted to match the NHANES data. Two different groups in the CAESAR sample, obtained from North America and the Netherlands were reweighted. Maximum difference was observed at the tails of the distribution as expected. In the reweighted Netherlands CAESAR data, stature varied up to 5 mm, mass up to 0.6 kg, and BMI up to 2.1 in the statistical measures considered here. Whereas in the reweighted North American CAESAR data, stature varied only up to 1 mm, mass up to 0.8 kg, and BMI up to only 0.3 in the statistical measures considered here. Therefore the North American CAESAR data represented the NHANES 2011-14 adults data very closely. The CAESAR data from Netherlands, although represented the NHANES data very well, was not as close as the reweighted North American CAESAR data. This could be because of the differences in racial composition between the US and Netherlands. Such demographic differences were relatively low between the US and North

Table 4.1: Statistics of CAESAR reweighted to represent NHANES 2011-14 population

					absolute difference from NHANES	
statistic	measure	NHANES 2011-14	weighted CAESAR N.America	weighted CAESAR Netherlands	weighted CAESAR N.America	weighted CAESAR Netherlands
5^{th} percentile	height(mm)	1534	1535	1538	1	4
	weight (kg)	53.9	53.7	53.4	0.2	0.5
	BMI	20.0	20.0	20.0	0.0	0.0
50^{th} percentile	height(mm)	1689	1688	1692	1	3
	weight (kg)	79.6	79.6	79.4	0.0	0.2
	BMI	27.6	27.6	27.4	0.0	0.2
95^{th} percentile	height(mm)	1859	1859	1864	0	5
	weight (kg)	123.0	122.2	122.4	0.8	0.6
	BMI	42.5	42.3	40.4	0.2	2.1
mean	height(mm)	1693	1694	1696	1	4
	weight (kg)	83.0	82.9	82.7	0.1	0.3
	BMI	28.9	28.8	28.6	0.1	0.3
standard deviation	height(mm)	100	99	99	1	1
	weight (kg)	22.0	21.6	21.0	0.4	1.0
	BMI	7.1	6.8	6.5	0.3	0.6

American populations.

Table 4.2 compares ANSUR data and North American CAESAR data reweighted to match the AIST data collected in Japan using statistical measures described earlier. The North American CAESAR data was observed to be closer than ANSUR data in representing the Japanese population upon reweighting. One of the reasons for this might have been the better fitness of the army population in ANSUR data compared to the civilian population. The maximum difference between the reweighted North American CAESAR and original Japanese data was 4 mm in stature, 0.4 kg in mass, and 0.3 in BMI, among the statistical measures considered here. These difference are very low and hence the reweighted data can be used to design for the Japanese population, thereby providing detailed 3D scans from the CAESAR database.

For the third reweighting test, both North American and Netherlands CAESAR data were reweighted to represent the ANSUR population. Table 4.3 gives the differences in the statistical measures described earlier, between the reweighted datasets and the original ANSUR data. It was observed that the maximum difference among these statistical measures was only 2 mm in stature, 0.2 kg in mass, and 0.1 in BMI, for the reweighted North American CAESAR population. This better represented the ANSUR population, than the reweighted Netherlands CAESAR data. This is expected from the closer demographics of the North American CAESAR to the ANSUR population compared to that of Netherlands CAESAR data.

These comparisons of stature, mass and BMI of reweighted data with their original data

Table 4.2: Statistics of ANSUR and CAESAR reweighted to represent Japanese population

statistic	measure	Japanese data	weighted ANSUR	weighted CAESAR N America	absolute difference from Japanese data: weighted ANSUR	absolute difference from Japanese data: weighted CAESAR N America
5th percentile	height (mm)	1465	1477	1468	12	4
	weight (kg)	44.8	45.5	45.0	0.7	0.2
	BMI	18.0	18.4	18.2	0.4	0.2
50th percentile	height (mm)	1627	1627	1625	1	2
	weight (kg)	56.4	56.1	56.7	0.3	0.3
	BMI	21.3	21.3	21.3	0.0	0.0
95th percentile	height (mm)	1789	1785	1785	4	4
	weight (kg)	73.0	73.2	73.3	0.1	0.3
	BMI	26.1	26.0	26.2	0.2	0.0
mean	height (mm)	1631	1632	1632	1	1
	weight (kg)	57.3	57.7	57.7	0.3	0.4
	BMI	21.5	21.6	21.6	0.1	0.1
standard deviation	height (mm)	97	93	98	4	1
	weight (kg)	9.1	9.0	9.5	0.0	0.4
	BMI	2.5	2.4	2.9	0.1	0.3

provides a means to evaluate the effectiveness of the reweighting process. Comparing the different statistical measures of stature, mass, and BMI between the reweighted data and actual data gives designers more surety on using other anthropometric data proportional to these measures, from the reweighted data. Comparing table 4.1 with table 3.2 shows that the differences in the compared statistics of reweighted data from the actual data, obtained with the new method was far lower than any of the existing methods.

Table 4.3: Statistics of CAESAR reweighted to represent ANSUR population

					absolute difference from ANSUR	
statistic	measure	ANSUR	weighted CAESAR N America	weighted CAESAR Netherlands	weighted CAESAR N America	weighted CAESAR Netherlands
5^{th} percentile	height (mm)	1548	1548	1549	0	1
	weight (kg)	52.5	52.6	52.5	0.1	0.0
	BMI	20.1	20.1	20.3	0.0	0.1
50^{th} percentile	height (mm)	1682	1683	1680	1	2
	weight (kg)	68.9	68.9	68.6	0.0	0.3
	BMI	24.5	24.5	24.4	0.0	0.1
95^{th} percentile	height (mm)	1838	1840	1842	2	4
	weight (kg)	93.7	93.9	93.4	0.2	0.3
	BMI	30.1	30.1	30.2	0.1	0.1
mean	height (mm)	1686	1686	1686	0	1
	weight (kg)	70.7	70.7	70.7	0.0	0.0
	BMI	24.7	24.7	24.7	0.0	0.0
standard deviation	height (mm)	90	91	91	0	1
	weight (kg)	12.9	13.0	13.0	0.1	0.1
	BMI	3.1	3.1	3.1	0.1	0.1

Chapter 5

Evaluation of new method using multivariate design test cases

Until now the reweighted data was compared with that of the original weighted population using descriptive statistics. These included mean, standard deviation and different percentile values. But it is important to look at the effectiveness of the reweighting technique in real-world applications. Design applications are rarely limited to univariate problems where only one design variable needs to be decided at a time. Most design problems are multivariate, where several design variable together decide the accommodation level of the product. For example, in case of the design of a chair, a person whose hip breadth is less than the seat width is still disaccommodated if his/her buttock-popliteal length is less than the seat depth. Therefore, it is necessary calculate accommodation and decide design variables by considering multiple dimensions simultaneously.

A suite of multivariate test cases suggested by G. Nadadur and M. Parkinson [1] were used to evaluate the effectiveness of the reweighted data in such design problems. These problems comprise of commonly-used and widely-available body dimensions. The following study makes use of these test cases to demonstrate the use of the new reweighting technique in reweighting detailed anthropometric data for use in designing products for a different population. Only those target populations which has detailed anthropometric data collected from surveys are chosen as this allows to compare the accommodation level of the resulting product calculated using the surveyed data of the target population and the different data reweighted for the target population. Hence the error in the perceived accommodation level of the designed product with the reweighted data can be estimated, in real-world design situations.

5.1 Virtual fitting trials

Virtual fitting trials aim to find the number of individuals from a dataset that are accommodated by a particular design when multiple design variables are considered simultaneously. An individual is considered accommodated in the design only if all the design variables are able to accommodate the corresponding body measurements of the individual under consideration. This largely decreases the number of accommodated individuals compared to univariate analysis where only one design variable is considered at a time to check the accommodation level. In virtual fitting trails, values have to be chosen for each of the design dimensions and then each individual in the dataset is compared to check whether that person is accommodated with the selected dimensions. The total accommodation is then calculated based on the number of individuals considered accommodated and the total number of records in that dataset.

In order to attain a certain level of accommodation for a particular design, several different combinations of values for the design dimensions are possible. For this test, the design dimensions were randomly chosen from either the 1-10 and 90-99 percentile ranges of the corresponding body measures in the original representative data, or from 1-5 and 95-99 percentile ranges of the same. The following sections describe the product dimensions and corresponding body measures relevant to each of the design cases. The accommodation level is calculated with the original data representative of an actual population, a second dataset reweighted to match this representative data, and the unweighted second dataset. It is essential that all the datasets used contain the necessary body measures required in each of the design cases. Since the design dimensions are chosen from particular percentiles of the available anthropometric data, the difference in calculated accommodation level would drive the choice of design dimensions. Therefore, if the reweighted data was used to decide the product dimensions, the closer the calculated accommodation of the reweighted data is to that of the original representative data, the more accurate the chosen design dimensions would be to the values required for the intended accommodation level.

5.2 Design of cab geometry of work equipment

The first case involves the cab geometry of work equipment. This case would apply to work equipment such as fork lifts and cranes where the operator has to be comfortable for long periods of work hours as well as be able to effectively reach out to all the controls in the cabin. Table 5.1 gives the product dimensions that need to be defined, their corresponding anthropometric measure that drive the design values, as well as their relationships for cab geometry design. Figure 5.1 shows the dimensions of cab geometry guided by each of the body measures. In order to demonstrate the use of reweighted anthropometric data for design of the cab geometry, North American CAESAR data was reweighted to represent the ANSUR data.

Virtual fitting tests as described in the previous section were conducted with the four body measures mentioned in table 5.1. For cases with B≤P (bideltoid breadth, foot length, sitting

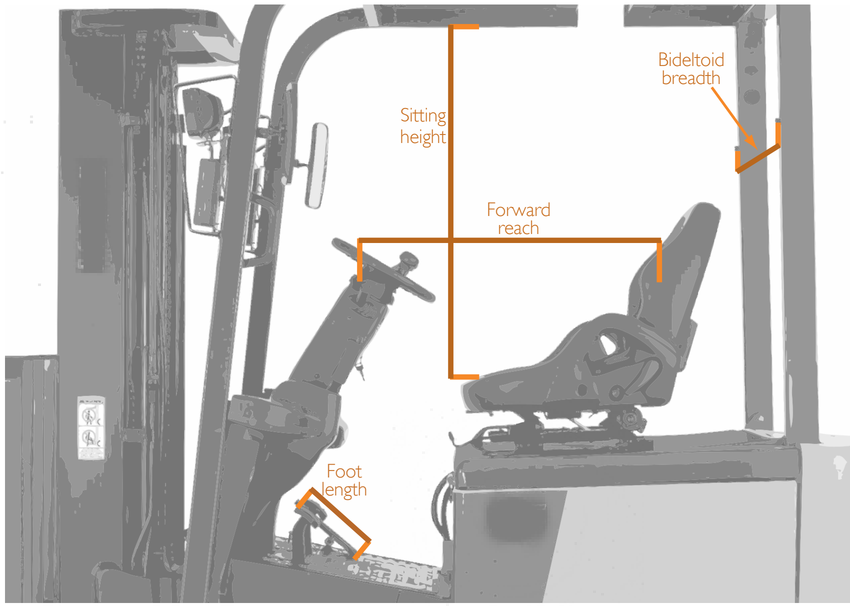

Figure 5.1: Body measures that guide the design of cab geometry. Adapted from [1].

Table 5.1: Product dimensions, relevant body measures, and design constrains involved in design of cab geometry of work equipment. Adapted from [1].

Product dimensions (P)	Body measure (B)	Suggested design constrain
Cab width	Bideltoid breadth	B≤P
Foot pedal length	Foot length	B≤P
Control distance	Forward reach	B≥P
Cab height	Sitting height	B≤P

height), P was chosen from the higher tail of the distribution, that is from either the percentile range 90-99 or 95-99. For cases with B≥P (forward reach), P was chosen from the lower tail of the distribution, that is from either the percentile range 1-10 or 1-5. The difference in accommodation attained with the CAESAR data improved from 6.4% to 1.0% upon weighting for percentile values of 1-10 and 90-99. Whereas for percentile values of 1-5 and 95-99, it improved from 6.0% to 0.8%. Therefore it was observed that reweighting the data before using it in the actual design process made the perceived accommodation more accurate. This in turn affects the choice of design dimensions. The design dimensions would have otherwise been made unnecessarily large or small thereby increasing the costs or adversely affecting accommodation level of the design. The difference of 0.8 to 1.0% in accommodation levels signify the range of error that may be observed in the accommodation level of the cab geometry when using reweighted data instead of real surveyed data for the target population to make the design decisions.

Table 5.2: Mean accommodation of the cab geometry of work equipment with its dimensions picked randomly from given range of percentiles of ANSUR data, when compared with ANSUR data, CAESAR data reweighted with Japanese data and unweighted CAESAR data

percentile ranges	data	accommodation (%) mean	standard deviation	mean absolute difference from ANSUR
1-10 & 90-99	ANSUR	82.1	3.9	
	CAESAR (weighted)	82.8	4.2	1.0
	CAESAR (unweighted)	75.8	4.1	6.4
1-5 & 95-99	ANSUR	89.8	1.8	
	CAESAR (weighted)	90.6	1.8	0.8
	CAESAR (unweighted)	83.8	2.0	6.0

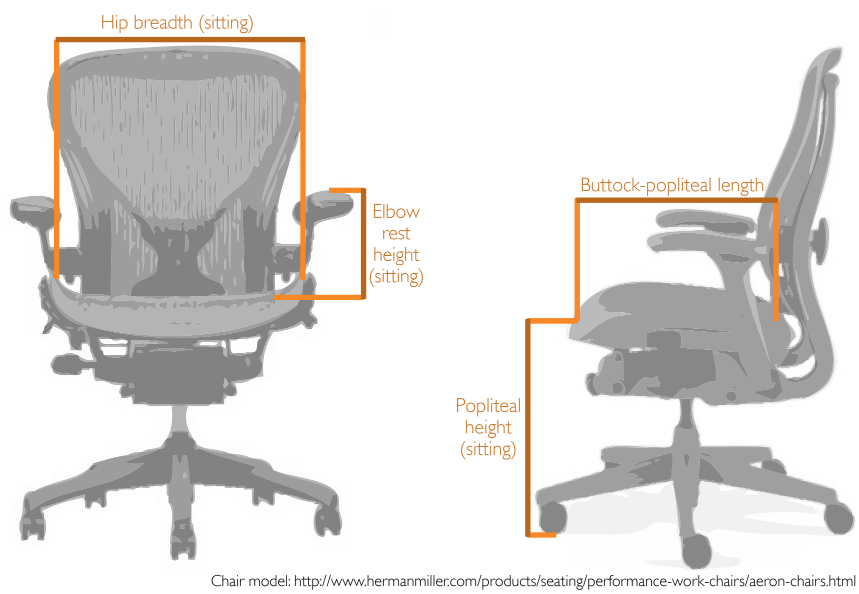

Figure 5.2: Body measures that guide the design of office chair. Adapted from [1].

5.3 Design of office chair

The second case involves the design of office chairs. Table 5.3 gives the product dimensions that need to be defined, their corresponding anthropometric measure that drive the design values, as well as their relationships for the chair design. Figure 5.2 shows the dimensions of chair guided by each of the body measures. In this case, ANSUR data was reweighted to represent the Japanese population.

Virtual fitting tests as described in section 5.1 were conducted with the four body measures mentioned in table 5.3. For cases with B≤P (elbow rest height and hip breadth), P was chosen from the higher tail of the distribution, that is from either the percentile range 90-99 or 95-99.

Table 5.3: Product dimensions, relevant body measures, and design constrains involved in design of office chairs. Adapted from [1].

Product dimensions (P)	Body measure (B)	Suggested design constrain
Seat depth	Butt-popliteal length	B≥P
Arm rest height	Elbow rest height (sit)	B≤P
Seat width	Hip breadth (sit)	B≤P
Seat height	Popliteal height (sit)	B≥P

Table 5.4: Mean accommodation of the seat design with its dimensions picked randomly from given range of percentiles of Japanese anthropometric data, when compared with Japanese data, ANSUR data reweighted with Japanese data and unweighted ANSUR data

		accommodation (%)		
percentile ranges	data	mean	standard deviation	mean absolute difference from Japanese data
1-10 & 90-99	Japanese	81.7	3.9	
	ANSUR (weighted)	87.5	3.7	5.9
	ANSUR (unweighted)	61.9	7.9	19.8
1-5 & 95-99	Japanese	89.6	1.7	
	ANSUR (weighted)	91.8	0.9	2.2
	ANSUR (unweighted)	71.7	3.6	17.9

For cases with B≥P (buttock-popliteal length and popliteal height), P was chosen from the lower tail of the distribution, that is from either the percentile range 1-10 or 1-5. The difference in accommodation attained with the ANSUR data improved from 19.8% to 5.9% upon weighting for percentile values of 1-10 and 90-99. Whereas for percentile values of 1-5 and 95-99, it improved from 17.9% to 2.2%. Therefore it was observed that reweighting the data before using it in the design process made the perceived accommodation much more accurate. The difference of 2.2 to 5.9% is not negligible. This error could primarily be because of the difference in racial composition between the two populations as well as the difference between civilian and army populations.

5.4 Design of crutch

The third case involves the design of crutch. The grip circumference and support-grip length of the crutch were identified as dimensions that define the accommodation level of the product as illustrated in figure 5.3. Table 5.5 gives the product dimensions that need to be defined, their corresponding anthropometric measure that drive the design values, as well as their relationships for the chair design. In this case, CAESAR data was reweighted to represent the Japanese population.

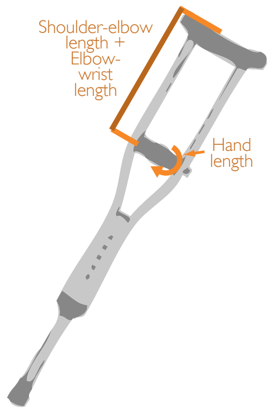

Figure 5.3: Body measures that guide the design of crutch. Adapted from [1].

Table 5.5: Product dimensions, relevant body measures, and design constrains involved in design of crutches. Adapted from [1].

Product dimensions (P)	Body measure (B)	Suggested design constrain
Grip circumference	Hand length	B≥P
Support-grip length	Elbow-wrist height + Shoulder-elbow length	B≥P

Virtual fitting tests as described in section 5.1 were conducted with the four body measures mentioned in table 5.3. Both the body measures (hand length and elbow-wrist height + shoulder-elbow length) were chosen from the lower tail of the distribution, that is from either the percentile range 1-10 or 1-5. This was because both the body measures had to be greater than or equal to their corresponding product dimensions, as given in table 5.5. The difference in accommodation attained with the CAESAR data improved from 5.8% to 1.9% upon weighting for percentile values of 1-10 and 90-99. Whereas for percentile values of 1-5 and 95-99, it improved from 3.5% to 1.0%. Therefore it was observed that reweighting the data before using it in the design process made the perceived accommodation more accurate. The difference of 1.0 to 1.9% is small and therefore would not considerably affect the design. This error could primarily be because of the difference in racial composition between the two populations.

Table 5.6: Mean accommodation of the crutch design with its dimensions picked randomly from given range of percentiles of ANSUR data, when compared with ANSUR data, Japanese anthropometric data reweighted with ANSUR data and unweighted Japanese data

		accommodation (%)		
percentile ranges	data	mean	standard deviation	mean absolute difference from Japanese data
1-10 & 90-99	Japanese	91.6	2.7	
	CAESAR (weighted)	92.8	2.7	1.9
	CAESAR (unweighted)	97.4	1.3	5.8
1-5 & 95-99	Japanese	94.5	1.0	
	CAESAR (weighted)	94.9	1.4	1.0
	CAESAR (unweighted)	98.5	0.6	3.5

Chapter 6

Conclusion

Unavailability of detailed anthropometric data causes products to be designed for populations different from their target populations. The available detailed anthropometric data can be reweighted to represent the target population using one of the existing reweighting techniques. But the three existing methods were shown to have reweighted data with a large error particularly in the tails of the distribution. Most of the design decisions are also made based on the body dimensions in the tails of the population distribution. Therefore, a new method to reweight anthropometric data is demonstrated to obtain detailed anthropometric data for the target population with much lower and acceptable errors. This method can be used even when only two out of the three measures, stature, mass, and BMI, available for a target population.

The number of bins used in the weighting for male CAESAR data using male NHANES 2011-14 data was 993. This is more than 220% of the number of bins used in Harrison method while still using less number of predictors, and more than 620% of the number of bins used in Paquette method. In all the existing methods, the number of bins are fixed. But the number of bins formed in the new weighting technique varies according to the datasets used, to provide better reweighting.

Most of the issues with reweighting data arose not from the technique itself but from the lack of a large sample size for the data. This was the case with reweighting for the Japanese population for which a limited sample size was available. But since the reweighting requires only any two of stature, mass and BMI, it is much easier to collect data with a large sample size. The cost of such a survey would also be much less than a survey to collect detailed anthropometric data. Racial composition was another factor which caused increased errors in the reweighted data. If detailed anthropometric data can be obtained for a particular population, it can be reweighted reliably for all populations with similar racial composition. Inclusion of race bins was found to increase the error when reweighting NHANES data. But this can be attributed to the limited data on minority races. If the sample size is large enough and there is a broad range of data available for each race, the target population need not be similar in racial composition either, as the binning strategy can then incorporate race bins and still provide better accuracy.

Reweighting data for several different populations were demonstrated, namely US civilian population, US army population and Japanese population. Data from these three populations as well as CAESAR data collected from North America and the Netherlands were used for this. The results showed that the error in percentiles of stature, mass and BMI were very low when the right data was chosen for reweighting. This was again dependent on the similarity in race and occupational groups. The main application for reweighting is in design of products. This was demonstrated through three multivariate design cases. The accommodation level of the designed product calculated using the reweighted data and the actual data for the target population were very close. This shows that the reweighted data may be used to decide the design dimensions to attain the desired accommodation level.

The principal contribution of this thesis is that it provides a method to reweight anthropometric data to represent a target population with acceptable errors. This was not achieved with the existing reweighting methods. Using relevant anthropometric data in design helps to counter many of the health, safety and performance issues associated with the product, described in Chapter 1. Sections 5.2 and 5.3 demonstrate the use of this method in obtaining detailed anthropometric data for design of truck cab geometry and chairs. It is demonstrated that the error in perceived accommodation of the product is reduced by a large margin, when weighted anthropometric data which better represents the target population is used in place of unweighted data. This improves the sizing of the resulting products as designers can pick product dimensions based on the accommodation level calculated for that particular set of product dimensions using this reweighted anthropometric data which better represents the target population. The method presented in this thesis was found to be an effective way to assign weights to unweighted data by comparing only the basic predictors of stature, mass and BMI. The method presented here was also shown to model the target population more closely than the existing method. This provides the potential to use detailed anthropometric data in designing products and environments more accurately for target populations for which, only limited data such as stature and mass are available. It also allows more anthropometric measurement records to be added to existing database with re-weighting without affecting the population statistics. This technique can also be applied in many other fields which requires weighting of data, not limited to anthropometric data.

Although the reweighted data can be verified for accuracy by comparing several statistics of stature, mass and BMI, this can only verify the reweighted data up to an extend. The low errors in these three measures could signify low errors in anthropometric measures that are proportional to one or more of these three basic measures. Buttock-popliteal distance and knee height are examples for such anthropometric measures. But in case of anthropometric measures that are not proportionate with any of these three basic measures, it is impossible to verify the accuracy of the reweighted data. An example for this is the interpupillary distance which is the distance between the center of the pupils of the two eyes of a person. Future research may be conducted on reweighting the available data on such body measures to represent those of the target population, and verifying it.

Appendix

Documentation of data reweighting

A.1 Data reweighting MATLAB script

```
load Nm1114.txt
%[Nm 1=ht 2=wt 3=bmi 4=ethnic 5=gender 6=age
%7=st.wt +8=bin#HtWt 9=bin#age], ages 18-65

load Cm.txt
%[Cm 1=ht 2=wt 3=bmi 4=ethnic 5=gender 6=age
%+7=st.wt 8=bin# 9=bin#valid]

pl=zeros(99,9);
error=zeros(3,3);
tic
Nm=Nm1114;
Nm(:,1) = Nm(:,1)*10;
Cm(:,8) = 1:size(Cm,1);    %assign bin numbers to each data pt
Cm(:,9) = 0;

%Ranges of stature, mass and BMI in dataset for use in
%the function for Euclidean distance.
r1=max(Nm(:,1))-min(Nm(:,1));
r2=max(Nm(:,2))-min(Nm(:,2));
r3=max(Nm(:,3))-min(Nm(:,3));

%find closest C point for every N point and mark C point as valid
```

```
for i=1:size(Nm,1)
    [Nm(i,8)]=closestNbin(Nm(i,1:6),Cm);
    Cm(Nm(i,8),9)=1;
end
%find closest N point for every invalid C point and copy bin#
for i=1:size(Cm,1)
    if (Cm(i,9)==0)
        [Cm(i,8)] = closestNbin(Cm(i,1:6),Nm);
    end
end
%create table of all valid bin#, count #Cm in each bin
Bins=zeros(1,3); %[Bins 1=bin# 2=sum_of_st.wt 3=#Cm]
for i=1:size(Cm,1)
    if (Cm(i,9)==1)
        Bins(end+1,1)=Cm(i,8);
        for j=1:size(Cm,1)
            if (Cm(j,8)==Cm(i,8))
                Bins(end,3) = Bins(end,3) + 1;
            end
        end
    end
end
Bins = Bins(2:end,:);
%sum of st.wts from N for each bin
for i=1:size(Bins,1)
    for j=1:size(Nm,1)
        if(Bins(i,1)==Nm(j,8))
            Bins(i,2) = Bins(i,2)+Nm(j,7);
        end
    end
end
%Assign st.wt. to C using by matchin bin# with Bins table
for i=1:size(Cm,1)
    for j=1:size(Bins,1)
        if (Cm(i,8)==Bins(j,1))
            Cm(i,7)=Bins(j,2)/Bins(j,3);
        end
    end
end
% percentiles {1,2,3=Nm[ht,wt,bmi], 4,5,6=Cm[ht,wt,bmi],
```

```
% 7,8,9=error[ht,wt,bmi]}
for i=1:3
    pl(:,i)=wprctile(Nm(:,i),1:99,Nm(:,7));
    pl(:,i+3)=wprctile(Cm(:,i),1:99,Cm(:,7));
    pl(:,i+6) = abs( pl(:,i)-pl(:,i+3) );
end
% error [i=1:10000; j=mean,sd,max; k=ht,wt,bmi]
for i=1:3
    error(1,i) = mean( pl(:,i+6) );
    error(2,i) = std( pl(:,i+6) );
    error(3,i) = max( pl(:,i+6) );
end
```

A.2 MATLAB script for function to find closest datapoint from the second dataset for a particular datapoint

```
function [bin]= closestNbin(x,L)
dist=999999;
bin=0;
for i=1:size(L,1)
    dist2 = sqrt( ((x(1)-L(i,1))/r1)^2 + ((x(2)-L(i,2))/r2)^2
                                       + ((x(3)-L(i,3))/r3)^2 );
    if (dist2<dist)
        dist = dist2;
        bin = L(i,9);
    end
    % To exclude points which are too far
    if (dist > sqrt( (50/r1)^2 + (5/r2)^2 + (2/r3)^2 ))
        bin = 0;
    end
end
```

A.3 MATLAB script for Monte Carlo simulation to find error when using different sample sizes

```
load Nm1114.txt
%[1=ht 2=wt 3=bmi 4=ethnic 5=gender 6=age +7=st.wt 8=bin# 9=bin#valid]
load Nm0710.txt
res=table;
```

```
parfor k=1:20
    pl=zeros(99,9);
    error=zeros(30,3,3);
    tempres=zeros(3,3);
    for m=1:30
        Nm=Nm1114;
        RIx = randsample(1:size(Nm0710,1),k*100);
        Cm = Nm0710(RIx,1:6);
        Cm(:,8) = 1:size(Cm,1);    %assign bin numbers to each data pt
        Cm(:,9) = 0;
        %find closest C point for every N point and mark C point as valid
        for i=1:size(Nm,1)
            [Nm(i,8)]=closestNbin(Nm(i,1:6),Cm);
            Cm(Nm(i,8),9)=1;
        end
        %find closest N point for every invalid C point and copy bin#
        for i=1:size(Cm,1)
            if (Cm(i,9)==0)
                [Cm(i,8)] = closestNbin(Cm(i,1:6),Nm);
            end
        end
        %create table of all valid bin#, count #Cm in each bin
        Bins=zeros(1,3); %[Bins 1=bin# 2=sum_of_st.wt 3=#Cm]
        for i=1:size(Cm,1)
            if (Cm(i,9)==1)
                Bins(end+1,1)=Cm(i,8);
                for j=1:size(Cm,1)
                    if (Cm(j,8)==Cm(i,8))
                        Bins(end,3) = Bins(end,3) + 1;
                    end
                end
            end
        end
        Bins = Bins(2:end,:);
        %sum of st.wts from N for each bin
        for i=1:size(Bins,1)
            for j=1:size(Nm,1)
                if(Bins(i,1)==Nm(j,8))
                    Bins(i,2) = Bins(i,2)+Nm(j,7);
                end
```

```
            end
        end
        %Assign st.wt. to C using by matchin bin# with Bins table
        for i=1:size(Cm,1)
            for j=1:size(Bins,1)
                if (Cm(i,8)==Bins(j,1))
                    Cm(i,7)=Bins(j,2)/Bins(j,3);
                end
            end
        end
        % percentiles {1,2,3=Nm[ht,wt,bmi], 4,5,6=Cm[ht,wt,bmi],
        % 7,8,9=error[ht,wt,bmi]}
        for i=1:3
            pl(:,i)=wprctile(Nm(:,i),1:99,Nm(:,7));
            pl(:,i+3)=wprctile(Cm(:,i),1:99,Cm(:,7));
            pl(:,i+6) = abs( pl(:,i)-pl(:,i+3) );
        end
        % error [i=1:10000; j=mean,sd,max; k=ht,wt,bmi]
        for i=1:3
            error(m,1,i) = mean( pl(:,i+6) );
            error(m,2,i) = std( pl(:,i+6) );
            error(m,3,i) = max( pl(:,i+6) );
        end
    end
    % results for graph [k=1:20; j=mean,std,max; i=ht,wt,bmi]
    for i=1:3
        for j=1:3
            %res(k,j,i) = mean( error(:,j,i) );
            tempres(j,i) = mean ( error(:,j,i) );

        end
    end
    resNew = table (k,tempres(1,1),tempres(1,2),tempres(1,3),
                     tempres(2,1),tempres(2,2),tempres(2,3),
                         tempres(3,1),tempres(3,2),tempres(3,3));
    res=[res;resNew];
    k
end
toc
```

A.4 MATLAB script to reweight data using Harrison method

```
load Nm1114_AF.txt
%[Cm 1=ht 2=wt 3=bmi 4=ethnic 5=gender 6=age +7=st.wt 8=bin#HtWt
      %9=bin#age]
load Nf1114_AF.txt
load Cm_EthSexAge.txt
load Cw_EthSexAge.txt
pl=zeros(99,9);
error=zeros(3,3);
Nm = vertcat(Nm1114_AF,Nf1114_AF);
Cm = vertcat(Cm_EthSexAge,Cw_EthSexAge);
Nm(:,1) = Nm(:,1)*10;
Nm(:,3) = Nm(:,2)./((Nm(:,1)/1000).*(Nm(:,1)/1000));
Cm(:,3) = Cm(:,2)./((Cm(:,1)/1000).*(Cm(:,1)/1000));
for i=1:3
    pl(:,i)=wprctile(Nm(:,i),1:99,Nm(:,7));
end
for i=1:size(Nm,1)
    if (Nm(i,1)>=pl(28,1) && Nm(i,1)<=pl(72,1) && Nm(i,2)>=pl(28,2)
                                               && Nm(i,2)<=pl(72,2))
        Nm(i,8)=5;
    elseif (Nm(i,1)<=pl(50,1) && Nm(i,2)>=pl(50,2))
        Nm(i,8)=1;
    elseif (Nm(i,1)>=pl(50,1) && Nm(i,2)>=pl(50,2))
        Nm(i,8)=2;
    elseif (Nm(i,1)<=pl(50,1) && Nm(i,2)<=pl(50,2))
        Nm(i,8)=3;
    elseif (Nm(i,1)>=pl(50,1) && Nm(i,2)<=pl(50,2))
        Nm(i,8)=4;
    end
    if (Nm(i,4)~=1 && Nm(i,4)~=2)
        Nm(i,4)=3;
    end
    if (Nm(i,6)>=18 && Nm(i,6)<=29)
        Nm(i,9)=1;
    elseif (Nm(i,6)>=30 && Nm(i,6)<=44)
        Nm(i,9)=2;
    elseif (Nm(i,6)>=45 && Nm(i,6)<=65)
```

```
        Nm(i,9)=3;
    end
end
for i=1:size(Cm,1)
    if (Cm(i,1)>=pl(28,1) && Cm(i,1)<=pl(72,1) && Cm(i,2)>=pl(28,2)
                                                     && Cm(i,2)<=pl(72,2))
        Cm(i,8)=5;
    elseif (Cm(i,1)<=pl(50,1) && Cm(i,2)>=pl(50,2))
        Cm(i,8)=1;
    elseif (Cm(i,1)>=pl(50,1) && Cm(i,2)>=pl(50,2))
        Cm(i,8)=2;
    elseif (Cm(i,1)<=pl(50,1) && Cm(i,2)<=pl(50,2))
        Cm(i,8)=3;
    elseif (Cm(i,1)>=pl(50,1) && Cm(i,2)<=pl(50,2))
        Cm(i,8)=4;
    end

    if (Cm(i,6)>=18 && Cm(i,6)<=29)
        Cm(i,9)=1;
    elseif (Cm(i,6)>=30 && Cm(i,6)<=44)
        Cm(i,9)=2;
    elseif (Cm(i,6)>=45 && Cm(i,6)<=65)
        Cm(i,9)=3;
    end
end
%Table of Bins(HtWt1-5,gender1-2,ethnic1-3,age1-3,
%1=#Nm|2=sum(NmStWt)|3=#Cm)
Bins=zeros(5,2,3,3,3);
for i=1:size(Nm,1)
    Bins(Nm(i,8),Nm(i,5),Nm(i,4),Nm(i,9),1)
                        =Bins(Nm(i,8),Nm(i,5),Nm(i,4),Nm(i,9),1)+1;
    Bins(Nm(i,8),Nm(i,5),Nm(i,4),Nm(i,9),2)
                        =Bins(Nm(i,8),Nm(i,5),Nm(i,4),Nm(i,9),2)+Nm(i,7);
end

for i=1:size(Cm,1)
    Bins(Cm(i,8),Cm(i,5),Cm(i,4),Cm(i,9),3)
                            =Bins(Cm(i,8),Cm(i,5),Cm(i,4),Cm(i,9),3)+1;
end
for i=1:size(Cm,1)
```

```
    Cm(i,7) = Bins(Cm(i,8),Cm(i,5),Cm(i,4),Cm(i,9),2)
                            / Bins(Cm(i,8),Cm(i,5),Cm(i,4),Cm(i,9),3);
end

csvwrite('CWtAF.csv',Cm)
csvwrite('NWtAF.csv',Nm)
```

A.5 MATLAB script to reweight data using Hudson method

```
%[Nm 1=ht 2=wt 3=bmi 4=ethnic 5=gender 6=age 7=st.wt], ages 18-65
load Nm1114_Hudson.txt
%[Cm 1=ht 2=wt 3=bmi 4=ethnic 5=gender 6=age +7=st.wt], ages 18-65
Nm11=Nm1114_Hudson;
load Nf1114_Hudson.txt
Nf11=Nf1114_Hudson;
load Cm_EthSexAge.txt
load Cw_EthSexAge.txt
pl=zeros(99,9);
error=zeros(3,3);

count=1;
%53 to 81 inches height, for the screening based on
%corresponding weight limit for each inch value of height
for i=53:81
    HtWt(count,1)=i;
    mx=0;
    for j=1:size(Nm11,1)
        if (round(Nm11(j,1)/2.54)==i)
            if (Nm11(j,2)>mx)
                mx=Nm11(j,2);
            end
        end
    end
    HtWt(count,2)=mx;
    mx=0;
    for j=1:size(Nf11,1)
        if (round(Nf11(j,1)/2.54)==i)
            if (Nf11(j,2)>mx)
```

```
                mx=Nf11(j,2);
            end
        end
    end
    HtWt(count,3)=mx;
    count=count+1;
end
males=0;
count=1;
for i=1:size(Cm_EthSexAge,1)
    for j=1:size(HtWt,1)
        if (round(Cm_EthSexAge(i,1)/25.4)==HtWt(j,1))
            if (Cm_EthSexAge(i,2) < (4.54+HtWt(j,2)))
                males(count)=i;
                count=count+1;
            end
        end
    end
end
females=0;
count=1;
for i=1:size(Cw_EthSexAge,1)
    for j=1:size(HtWt,1)
        if (round(Cw_EthSexAge(i,1)/25.4)==HtWt(j,1))
            if (Cw_EthSexAge(i,2) < (4.54+HtWt(j,3)))
                females(count)=i;
                count=count+1;
            end
        end
    end
end
Nm = vertcat(Nm11,Nf11);
Cm = vertcat(Cm_EthSexAge(males,:),Cw_EthSexAge(females,:));
Nm(:,1) = Nm(:,1)*10;
Nm(:,3) = Nm(:,2)./((Nm(:,1)/1000).*(Nm(:,1)/1000));
Cm(:,3) = Cm(:,2)./((Cm(:,1)/1000).*(Cm(:,1)/1000));

%Table of Bins(gender1-2,ethnic1-3,
%1=#Nm|2=sum(NmStWt)|3=#Cm)
Bins=zeros(2,3,3);
```

```
for i=1:size(Nm,1)
    Bins(Nm(i,5),Nm(i,4),1)=Bins(Nm(i,5),Nm(i,4),1)+1;
    Bins(Nm(i,5),Nm(i,4),2)=Bins(Nm(i,5),Nm(i,4),2)+Nm(i,7);
end

for i=1:size(Cm,1)
    Bins(Cm(i,5),Cm(i,4),3)=Bins(Cm(i,5),Cm(i,4),3)+1;
end
for i=1:size(Cm,1)
    Cm(i,7)=Bins(Cm(i,5),Cm(i,4),2) / Bins(Cm(i,5),Cm(i,4),3);
end

csvwrite('CWtHudson.csv',Cm)
csvwrite('NWtAF.csv',Nm)
```

A.6 MATLAB script to reweight data using Paquette method

```
load Nm1114_Paquette.txt
%[Cm 1=ht 2=wt 3=bmi 4=ethnic 5=gender 6=age +7=st.wt 8=bin#HtWt 9=bin#age]
Nm11=Nm1114_Paquette;
load Nf1114_Paquette.txt
Nf11=Nf1114_Paquette;
load Cm_EthSexAge.txt
load Cw_EthSexAge.txt
pl=zeros(99,9);
error=zeros(3,3);

Nm = vertcat(Nm11,Nf11);
Cm = vertcat(Cm_EthSexAge,Cw_EthSexAge);
Nm(:,1) = Nm(:,1)*10;
Nm(:,3) = Nm(:,2)./((Nm(:,1)/1000).*(Nm(:,1)/1000));
Cm(:,3) = Cm(:,2)./((Cm(:,1)/1000).*(Cm(:,1)/1000));

for i=1:3
    pl(:,i)=wprctile(Nm(:,i),1:99,Nm(:,7));
end
for i=1:size(Nm,1)
    if (Nm(i,6)<=20)
```

```
        Nm(i,9)=1;
    elseif (Nm(i,6)>=21 && Nm(i,6)<=25)
        Nm(i,9)=2;
    elseif (Nm(i,6)>=26 && Nm(i,6)<=35)
        Nm(i,9)=3;
    else
        Nm(i,9)=4;
    end
end
for i=1:size(Cm,1)
    if (Cm(i,6)<=20)
        Cm(i,9)=1;
    elseif (Cm(i,6)>=21 && Cm(i,6)<=25)
        Cm(i,9)=2;
    elseif (Cm(i,6)>=26 && Cm(i,6)<=35)
        Cm(i,9)=3;
    else
        Cm(i,9)=4;
    end
end
%Table of Bins(gender1-2,ethnic1-4,age1-4,
%1=#Nm|2=sum(NmStWt)|3=#Cm)
Bins=zeros(2,4,4,3);
for i=1:size(Nm,1)
    Bins(Nm(i,5),Nm(i,4),Nm(i,9),1)=Bins(Nm(i,5),Nm(i,4),Nm(i,9),1)+1;
    Bins(Nm(i,5),Nm(i,4),Nm(i,9),2)
                            =Bins(Nm(i,5),Nm(i,4),Nm(i,9),2)+Nm(i,7);
end
for i=1:size(Cm,1)
    Bins(Cm(i,5),Cm(i,4),Cm(i,9),3)=Bins(Cm(i,5),Cm(i,4),Cm(i,9),3)+1;
end
for i=1:size(Cm,1)
    Cm(i,7)=Bins(Cm(i,5),Cm(i,4),Cm(i,9),2) / Bins(Cm(i,5),Cm(i,4),Cm(i,9),3);
end

csvwrite('CWtPaquette.csv',Cm)
csvwrite('NWtAF.csv',Nm)
```

A.7 MATLAB script to find effect of increasing number of bins on reweighting data

```
tic
load Nm1114.txt
load Cm.txt
Nm1114(:,4)=Nm1114(:,7);
Nm=Nm1114(:,1:4);
%Convert cm to mm
Nm(:,1)=Nm(:,1)*10;
%Set max number of bins on each axis here
maxBins=10;
%These will be partitions for the 10 bins each along stature and mass axes
%1st value will be min(Nm)-some value, last will be max(Nm)+some value,
%-in between are percentiles
plSt=zeros(maxBins+1,1);
plMs=zeros(maxBins+1,1);
plSt(1)=min(Nm(:,1))-30;
plMs(1)=min(Nm(:,2))-3;
pl=zeros(99,9); %martix to store percentile values for results
%-matrix to store error's mean, sd, max
%error(binningdivisions, mean/sd/max, stature/mass/bmi
error=zeros(maxBins-1,3,3);
%This variable=0 if there are no bins in the end which didnt have at least
%-1 Nm and 1 Cm
invalid=zeros(maxBins-1,1);
%To record the number valid bins after merging, in each case
realBins=zeros(maxBins-1,1);
%Find the value of these partitions based on wtd percentiles
for i=1:(maxBins-1)
    for j=1:i
        plSt(j+1)=wprctile(Nm(:,1),j*(100/(i+1)),Nm(:,4));
        plMs(j+1)=wprctile(Nm(:,2),j*(100/(i+1)),Nm(:,4));
    end
    %Making last value as maximum
    plSt(i+2)=max(Nm(:,1))+30;
    plMs(i+2)=max(Nm(:,2))+3;
    %Decalre matrix to store no. of Nm (1st col), no. of Cm (2nd col) in
    %each bin
    %(stBin, msBin, 1-Sum of Nm wts & 2-No. of Nm & 3- No. of Cm in bin
```

```
% & 4-Bin no. & 5-Nth copy)
bins=zeros(i+1,i+1,5);
%Number these bins
for j=1:i+1
    for k=1:i+1
        bins(j,k,4)=((j-1)*(i+1))+k;
    end
end
%Assinging bin numbers to all data in two parts statureBin and massBin
for j=1:size(Nm,1)
    for stBin=1:i+1
        for msBin=1:i+1
            if( Nm(j,1)>plSt(stBin) && Nm(j,1)<=plSt(stBin+1)
                 && Nm(j,2)>plMs(msBin) && Nm(j,2)<=plMs(msBin+1) )
                Nm(j,5)=stBin;
                Nm(j,6)=msBin;
                %Add to sum of Nm weights in each bin
                bins(stBin,msBin,1)=bins(stBin,msBin,1)+Nm(j,4);
                %Add to no. of Nm weights in each bin
                bins(stBin,msBin,2)=bins(stBin,msBin,2)+1;
            end
        end
    end
end
for j=1:size(Cm,1)
    for stBin=1:i+1
        for msBin=1:i+1
            if( Cm(j,1)>plSt(stBin) && Cm(j,1)<=plSt(stBin+1)
                 && Cm(j,2)>plMs(msBin) && Cm(j,2)<=plMs(msBin+1) )
                Cm(j,5)=stBin;
                Cm(j,6)=msBin;
                %Add to no. of Cm weights in each bin
                bins(stBin,msBin,3)=bins(stBin,msBin,3)+1;
            end
        end
    end
end
%Check for bins which do not have at least 1 Nm and 1 Cm
%These bins should be merged with nearby bin which has 1 Nm and 1 Cm
%Loop it so that finally all bins will have 1 Nm and 1 Cm
```

```
for j=1:10
    for stBin=1:i+1
        for msBin=1:i+1
            if (bins(stBin,msBin,2)==0 || bins(stBin,msBin,3)==0)
                %invalid(i)=invalid(i)+1;
                for stepSt=[stBin,stBin-1,stBin+1]
                    for stepMs=[msBin,msBin-1,msBin+1]
                        if (stepSt==0 || stepSt==i+2)
                            stepSt=stBin;
                        end
                        if (stepMs==0 || stepMs==i+2)
                            stepMs=msBin;
                        end
                        if ( bins(stBin,msBin,5)==0
                                    || (bins(stBin,msBin,5)-1)
                                            >bins(stepSt,stepMs,5))
                            if ((bins(stBin,msBin,2)
                                        +bins(stepSt,stepMs,2))>0
                                        &&(bins(stBin,msBin,3)
                                        +bins(stepSt,stepMs,3))>0)
                                bins(stBin,msBin,4)
                                        =bins(stepSt,stepMs,4);
                                bins(stBin,msBin,5)
                                        =bins(stepSt,stepMs,5)+1;
                                for k=1:3
                                    bins(stBin,msBin,k)
                                        =bins(stBin,msBin,k)
                                            +bins(stepSt,stepMs,k);
                                    bins(stepSt,stepMs,k)
                                        =bins(stBin,msBin,k);
                                end
                            end
                        end
                    end
                end
            end
        end
    end
end
%Check no. of bins still without at least 1 Nm and 1 Cm
```

```
    for stBin=1:i+1
        for msBin=1:i+1
            if (bins(stBin,msBin,2)==0 || bins(stBin,msBin,3)==0)
                invalid(i)=invalid(i)+1;
            end
        end
    end
    %Distributing total wt of each bin among its Cm's
    for j=1:size(Cm,1)
        if ( Cm(j,5)*Cm(j,6) > 0 )
            Cm(j,4) = bins( Cm(j,5),Cm(j,6),1 ) / bins( Cm(j,5),Cm(j,6),3 );
        end
    end
    % percentiles {1,2,3=Nm[ht,wt,bmi], 4,5,6=Cm[ht,wt,bmi],
    % 7,8,9=error[ht,wt,bmi]}
    for j=1:3
        pl(:,j)=wprctile(Nm(:,j),1:99,Nm(:,4));
        pl(:,j+3)=wprctile(Cm(:,j),1:99,Cm(:,4));
        pl(:,j+6) = abs( pl(:,j)-pl(:,j+3) );
    end
    for j=1:3
        error(i,1,j) = mean( pl(:,j+6) );
        error(i,2,j) = std( pl(:,j+6) );
        error(i,3,j) = max( pl(:,j+6) );
    end
    realBins(i)=size(unique(bins(:,:,4)),1);
end
toc
% Plot and save graphs
x=realBins(:,1);
for i=1:3
    yMean=error(:,1,i);
    ySD=error(:,2,i);
    yMax=error(:,3,i);
    figure;
    %plot(x,yMean,'-',x,ySD,'-');
    plot(x,yMax,'-r');
    xlabel('Number of bins');
    switch i
        case 1
```

```
            texti = 'Height';
        case 2
            texti = 'Mass';
        case 3
            texti = 'BMI';
    end
    ylabel(texti);
    textname=['/Users/openlab/Documents/Box Sync/Documents/Matt/BinNumbers
                                    /Graphs/Max',texti,'Bins.pdf'];
    saveas(gcf,textname,'pdf');
    close(gcf)
end
```

Bibliography

[1] Nadadur, G. and M. B. Parkinson (2013) "A suite of multivariate test cases for anthropometry-based design," *Proceedings of the 2nd International Digital Human Modeling Symposium*.

[2] Garneau, C. J. and M. B. Parkinson (2016) "A survey of anthropometry and physical accommodation in ergonomics curricula," *Ergonomics*, **59**(1), pp. 143–154.

[3] Chaffin, D. B., G. Andersson, and B. J. Martin (1999) *Occupational biomechanics*, Wiley New York.

[4] Tilley, A. (1993) "The measure of man and woman: human factors in design," .

[5] Roe, R. (1993) *Automotive Ergonomics*, Taylor & Francis, London.

[6] HFES 300 Committee (2004) *Guidelines for Using Anthropometric Data in Product Design*, Human Factors and Ergonomics Society, Santa Monica, CA.

[7] Administration, U. S. N. H. T. S. and U. o. M. T. R. Institute (1985) *Anthropometry of Motor Vehicle Occupants: Procedures, summary findings and appendices*, Anthropometry of Motor Vehicle Occupants, The Administration.
URL `https://books.google.com/books?id=FqyXztipPAAC`

[8] BIFMA (2013) *G1 - Ergonomics Guideline for Furniture used in Office Work Spaces Designed for Computer Use*.

[9] Haslegrave, C. M. (1986) "Characterizing the anthropometric extremes of the population," *Ergonomics*, **29**(2), pp. 281–301.

[10] Hsiao, H., D. Long, and K. Snyder (2002) "Anthropometric differences among occupational groups," *Ergonomics*, **45**(2), pp. 136–152.

[11] Gordon, C. C., T. Churchill, C. E. Clauser, B. Bradtmiller, and J. T. McConville (1989) *Anthropometric survey of US army personnel: methods and summary statistics 1988*, *Tech. rep.*

[12] Hertzberg, H. T. E. (1972) "Engineering anthropology," *Human engineering guide to equipment design*, pp. 468–584.

[13] Paquette, S., C. Gordon, and B. Bradtmiller (2009) *Anthropometric survey (ANSUR) II pilot study: methods and summary statistics*, *Tech. rep.*

[14] Scott, P. A., P. D. Candler, and J.-C. Li (1996) "Stature and seat position as factors affecting fractionated response time in motor vehicle drivers," *Applied ergonomics*, **27**(6), pp. 411–416.

[15] Parkinson, M. B. and M. P. Reed (2006) "Improved head restraint design for safety and compliance," in *ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, pp. 133–139.

[16] Bittner, A. C. (1985) "Human factors engineering principles for minimizing adverse ship motion effects: Theory and practice," *Naval Engineers Journal*, **97**(4), pp. 205–213.

[17] Kennedy, R. S. and L. H. Frank (1985) *A review of motion sickness with special reference to simulator sickness*, *Tech. rep.*

[18] Johnson, W. H. and J. W. Mayne (1953) "Stimulus required to produce motion sickness; restriction of head movement as a preventive of airsickness; field studies on airborne troops." *The Journal of aviation medicine*, **24**(5), pp. 400–411.

[19] Lackner, J. R. and A. Graybiel (1984) "Elicitation of motion sickness by head movements in the microgravity phase of parabolic flight maneuvers." *Aviation, space, and environmental medicine*.

[20] Moroney, W. F. and M. J. Smith (1972) *Empirical Reduction in Potential User Population as the Result of Imposed Multivariate Anthropometric Limits.*, *Tech. rep.*

[21] Radjabi, A. (1983) "Carpets, hand-woven," *Encyclopaedia of occupational health and safety. 3rd ed. Geneva, Switzerland: International Labour Office*, **1**, pp. 416–418.

[22] Chaffin, D. B. (1987) "Occupational biomechanicsA basis for workplace design to prevent musculoskeletal injuries," *Ergonomics*, **30**(2), pp. 321–329.

[23] Easterby, R. (2012) *Anthropometry and biomechanics: theory and application*, vol. 16, Springer Science & Business Media.

[24] Lockett, J. F., R. Kozycki, C. C. Gordon, and E. Bellandi (2005) *Proposed integrated human figure modeling analysis approach for the army's future combat systems*, *Tech. rep.*

[25] U.S. Centers for Disease Control and Prevention (2016) *National Health and Nutrition Examination Survey, 2015-2016*, *Tech. rep.*, National Center for Health Statistics.

[26] Robinette, K. M., S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming (2002) *Civilian American and European Surface Anthropometry Resource (CAESAR), Final Report. Volume 1. Summary*, *Tech. rep.*

[27] Robinette, K. M., H. Daanen, and E. Paquet (1999) "The CAESAR project: a 3-D surface anthropometry survey," in *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*, IEEE, pp. 380–386.

[28] Robinette, K. M. (2000) "CAESAR measures up," *Ergonomics in design*, **8**(3), pp. 17–23.

[29] National Institute of Bioscience and Human-Technology (1996) "AIST Anthropometric Database 1991-1992, Human Body Dimensions Data for Ergonomic Design," *Nippon Shuppan Service, Tokyo.*

[30] Flannagan, C. A. C., M. A. Manary, L. W. Schneider, and M. P. Reed (1998), "An Improved Seating Accommodation Model with Application to Different User Populations," .
URL http://dx.doi.org/10.4271/980651

[31] You, H. and T. Ryu (2005) "Development of a hierarchical estimation method for anthropometric variables," *International journal of industrial ergonomics*, **35**(4), pp. 331–343.

[32] Nadadur, G. and M. B. Parkinson (2008) "Extrapolation of anthropometric measures to new populations," *SAE international journal of passenger cars-electronic and electrical systems*, **1**(2008-01-1858), pp. 567–573.

[33] Fromuth, R. C. and M. B. Parkinson (2008) "Predicting 5th and 95th percentile anthropometric segment lengths from population stature," in *ASME 2008 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, pp. 581–588.

[34] Garneau, C. J. and M. B. Parkinson (2009) "Including preference in anthropometry-driven models for design," *Journal of Mechanical Design*, **131**(10), p. 101006.

[35] Nadadur, G. and M. B. Parkinson (2013) "A Z-Score-Based Method to Synthesize Anthropometric Datasets for Global User Populations," in *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, pp. V03BT03A018–V03BT03A018.

[36] Hsiao, H., J. Whitestone, M. Wilbur, J. R. Lackore, and J. G. Routley (2015) "Seat and seatbelt accommodation in fire apparatus: anthropometric aspects," *Applied ergonomics*, **51**, pp. 137–151.

[37] Hsiao, H., J. Whitestone, T.-Y. Kau, and B. Hildreth (2015) "Firefighter Hand Anthropometry and Structural Glove Sizing A New Perspective," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **57**(8), pp. 1359–1377.

[38] Spasojević Brkić, V. K., Z. A. Veljković, T. Golubović, A. D. Brkić, and I. Kosić Šotić (2016) "Workspace design for crane cabins applying a combined traditional approach and the Taguchi method for design of experiments," *International Journal of Occupational Safety and Ergonomics*, **22**(2), pp. 228–240.

[39] Sanders, M. S. and B. E. Shaw (1985), "U.S. Truck Driver Anthropometric and Truck Work Space Data Survey: Sample Selection and Methodology," .
URL http://dx.doi.org/10.4271/852315

[40] Guan, J., H. Hsiao, B. Bradtmiller, T.-Y. Kau, M. R. Reed, S. K. Jahns, J. Loczi, H. L. Hardee, and D. P. T. Piamonte (2012) "US truck driver anthropometric study and multivariate anthropometric models for cab designs," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, p. 0018720812442685.

[41] Radja, G. A. (2012) *National automotive sampling systemcrashworthiness data system, 2011 analytical users manual*, *Tech. rep.*

[42] Harrison, C. R., K. M. Robinette, O. Air Force Research Laboratory (Wright-Patterson Air Force Base, H. E. Directorate., C. S. I. Division., U. States., A. F. M. Command., S. (Firm), N. O. v. T.-N. Onderzoek., and S. o. A. Engineers. (2002) *CAESAR : summary statistics for the adult population (ages 18-65) of the United States of America*, Air Force Research Laboratory, Human Effectiveness Directorate, Crew

System Interface Division, Air Force Materiel Command ; National Technical Information Service [distributor], Wright-Patterson AFB, Ohio; Springfield, Va.

[43] Hudson, J. A., G. F. Zehner, and K. M. Robinette (2003) *JSF CAESAR: construction of a 3-D anthropometric sample for design and sizing of joint strike fighter pilot clothing and protective equipment*, Tech. rep.