Classification of Body Fat based on Measurements and Body Composition*

Comparitive analysis of Body Fat classification based on BMI and Fat Percentage

Aamishi Avarsekar

December 14, 2024

BMI values fail to acknowledge the different body composition across both sexes. Using the Waist-Hip-Ratio is more accurate at estimating body fat levels using simple methods at home. This paper compares how this model performs against actual body composition such as isolated fat mass. Using this method makes self-identification of obsesity more accessible and preventive measures can be taken earlier.

Table of contents

---

*Code and data are available at: https://github.com/aamishi/ImprovedBodyFatClassification

1 Introduction

BMI is a common indicator for general health but it has its limitations. It does not take into consideration the actual body composition of an individual and other important factors such as sex and age. In its initial design, BMI was designed using only male subjects and thus cannot be applied to females. Due to its inability to capture body composition, a lot of false positives and false negatives are produced. Through this paper, I would like to compare how using the Waist-to-Hip Ratio (WHR) serves as a more accurate way of estimating your body fat levels at home. The inspiration for using the WHR stems from fat accumulation patterns and distribution of fat vertically. This also helps us further differentiate between closer classifications of BMI such as Underweight and Normal Categories and Overweight and Obese Categories. BMI is notorious for classifications. A popular example of this would be body builders being classified as Overweight according to BMI despite having lower levels for body fat and being at a lower risk of obesity health risks. On the other had, taller people are often excused as false negatives and are failed to be recognized as having higher body fat levels due to their height. The table Table 1 below includes the BMI categories as endorsed by the Government of Canada (Canada (2024)).

Table 1: BMI vs. Fat Mass by BMI Category

Table 1: BMI Categories as per the Government of Canada Guidelines

Category	BMI Range
Underweight	Below 18.5
Normal weight	18.5–24.9
Overweight	25.0–29.9
Obese - Class I	30.0–34.9
Obese - Class II	35.0–39.9
Obese - Class III	40.0 and above

The second prong of this paper is finding the relation between age, height and weight against actual body fat that is separate from a person's total body weight. This serves as a comparison between the classification of body fat levels based on physical measurements versus more anabolic storage of fat. Actual body fat is estimated using several methods such as skinfold calipers or DEXA scans (Kuzmar and Zalabata (2020)). Both of these methods require professionals to record and experts to decipher. It is unusual for common folk to opt for these methods without medication intervention or expert domain knowledge such as athletic coaches.

Through this paper, I analyse how common factors such as sex, height and weight compare against WHR and Body Scan classifications and how imporant are the main predictors for the respective models. I employ two multilevel logistical regression models to categorize people based primarily on the WHR and Body Fat Percentages. The paper follows a two pronged approach using two different data sets that estimate the classification in two separate ways. I then compare their accuracy against their respective data set.

It is a known fact that BMI generalizes several body types based only on height and weight considerations. However, through the work of this paper, it can be noted that High and Extreme levels of body fat are completely disregarded through WHR calculations. Both data sets and models show an increasing trend in body fat levels as weight increases. The model on measurements data predicts a smooth increase of body fat as a person's weight increases. However, the model based on body compositions predicts a steeper increase in body fat as a person's weight increases. This fact is more obvious when data is faceted by sex.

This paper uses R Core Team (2023), to simulate, download, clean and test the data, and create the model. More information on packages and specific methods is included later in the paper. This paper is structured as follows: The [Data Section]((**paper-data-section?**)) introduces the data sets, their variables and how they were tailored to obtain the needed information. The Model section discusses in greater detail the structure of both models and what information they convey. The results and analysis of the models and discussed in the

Results section. Lastly, the Discussion section discusses real-life implications from the findings of this paper, limitations faced and next steps.

2 Data Section

This paper uses R (R Core Team 2023), and the `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2023), `ggplot` (Wickham 2016), `arrow`(Richardson et al. (2024)), `dplyr` (Wickham et al. 2023) packages throughout the analysis to clean both data sets and create visualizations. `nnet` (Ripley 2023) is employed to fit and apply the models.

The primary motivator of this paper is the misclassification of diverse body types into the same BMI categories. The graph Figure 1 is a reference point for the main argument of the paper and shows how true body fat percentages differ within the same BMI category. As both data sets have different methods of estimating fitness levels, the aim of this paper is to generate a connection between the two data sets. In both data sets, I have created the variable `fat_percentage_category` that the model predict based on entirely different predictors measured in different scenarios. For each of the two data sets and their respective models, I have used the proxies, BMI and Body Fat Percentage, to determine the initial value for `fat_percentage_category`. The models then predict the new categories based on factors such as WHR and Body Fat Percentage. More discussion is included in the respective model section.

Figure 1

4
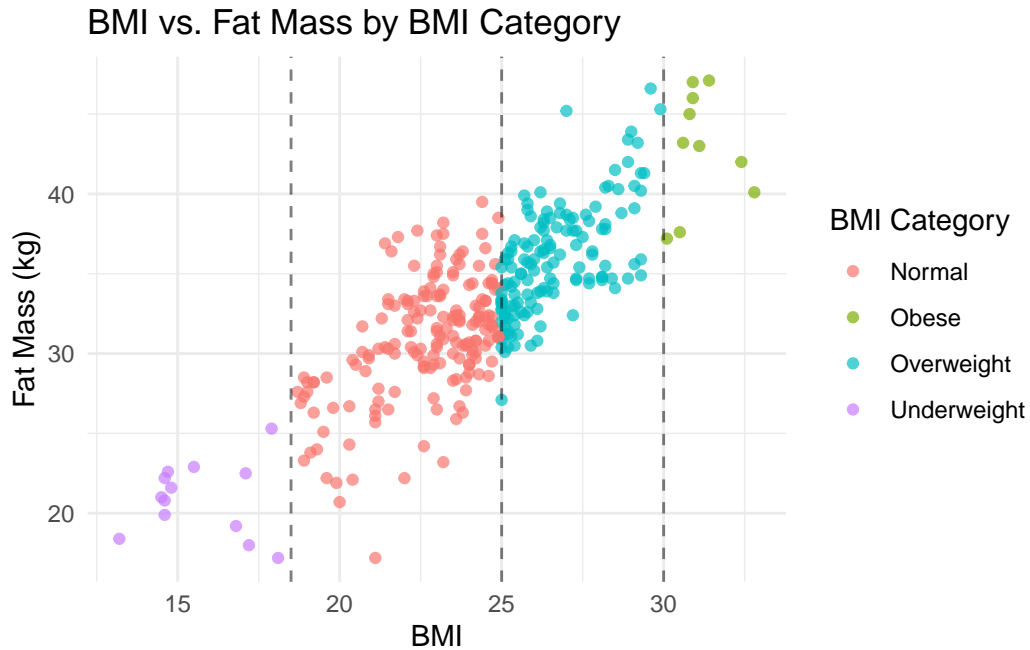
2.1 BodyM Data Set

This dataset, obtained from Amazon Web Services (AWS) (Amazon Web Services 2023), is used as training data for the model to predict body fat categories. It includes silhouettes of real test subjects to capture accurate body compositions. These silhouettes were collected to support the estimation of bodily measurements using Machine Learning techniques. However, the silhouettes are not used in this paper.

The BodyM Data Set, referred to as the Body Measurements data set throughout this paper, was collected by Ruiz et al. (Ruiz et al. 2022) with a focus on underrepresented body types in the estimation of fat and its subsequent health risks. The primary data captured in this collection consists of the front and lateral silhouettes of approximately [2000] test subjects. These silhouettes were then converted into black-and-white images for use in their augmentation model. The data set includes X male and Y female subjects, aged A to B. The body measurements in this data set were generated using their Adversarial Body Simulator (ABS), which was specifically designed to capture underrepresented body types. The S3 package (**reference?**) included three datasets: Training, Test A, and Test B. For the purposes of this paper, only the Training data set is used. The visual images of the test subjects were photographed and 3D-scanned by lab technicians.

2.1.1 The variables of my use:

For this paper, the BodyM Dataset is used to estimate the category of body fat that a person carries using circumference measurements. This means the circumferential length of body parts such as the waist and the hips. To build the model, I chose standard predictors such as gender, height and the weight of subjects as the base predictors. However, these predictors are not effective at differentiating body types. Therefore, to effectively construct the distribution of body fat in human body only through measurements, I created two variables called `waist_hip_ration` and `height_hip_ratio`. These variables are used to model the differences in the lateral and horizontal fat distributions. According to WHO (**reference?**), the waist-to-hip ratio, WHR, is a more accurate predictor of body fat around the abdominal area. Based on the study done in (**reference?**), the two sexes exhitbit difference tendenices for excess body fat accumulation, also known as the adipose fat. In males, excess fat tends to accumulate in the abdominal region, whereas in females, fat is more commonly stored in the hips. This ratio is indicative of fat distribution, and a higher WHR can be a precursing indicator of obesity (**reference?**). In the study done in [place], it is shown that an increase in height lowers the decreases the probability of being classified into a higher BMI category for both genders. Finally, to enhance the model's quality, ankle and wrist circumference measurements were also included. These measurements are particularly important for determining boundary values for fat classification (**reference?**). Excess fat accumulation at the wrists and ankles could indicate other health risks as seen in [].
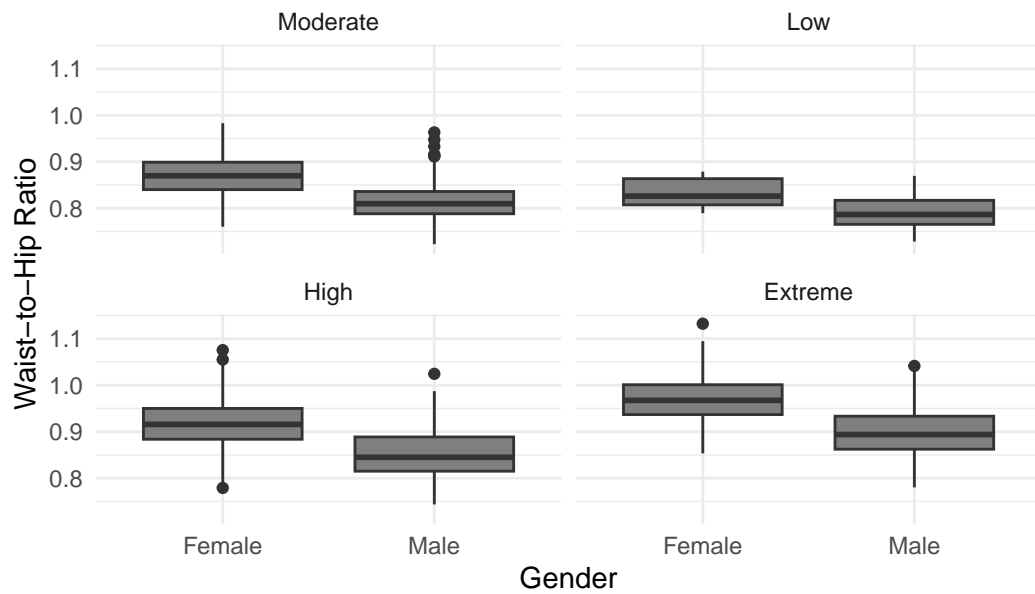
2.1.2 The estimand for the model

Using the above predictor variables, I contructed the variable `fat_percentage_category` as I recruit a multilevel linear regression model to classify the body fat. Initially, I wanted to predict the actual fat percentage, but this proved difficult as ... I decided to use a categorical variable as the estimand to produced useful comparisons against other categorical variables such as sex, height categories and weight categories. Additionally, `fat_percentage_category` was constructed using BMI as a proxy to gauge fat levels in a person. BMI is still widely used in medical scenarios to provide general diagnosis of health levels so I used that as a base to study the interactions between current systems and new offerings of WHR and height_hip_ratio. The relationship between `fat_percentage_category` and BMI was considered to be relatively straightforward to mimic the nature of the current applications of BMI. For BMI, I chose to consider the categories, `underweight`, `normal`, `overweight` and `obese`. These were calulated using the BMI formula: [insert BMI formula on a new line].

No difference between the genders was considered. Obesity categories using BMI consists of Obese I, Obese II and Obese III as they cause the increasing obesity relating such as diabetes, [ii] and [iii]. However, due to the lack of data points for Obesity II and III (BMI X - X), I chose to categorize all levels of obesity within the same `obesity` category.
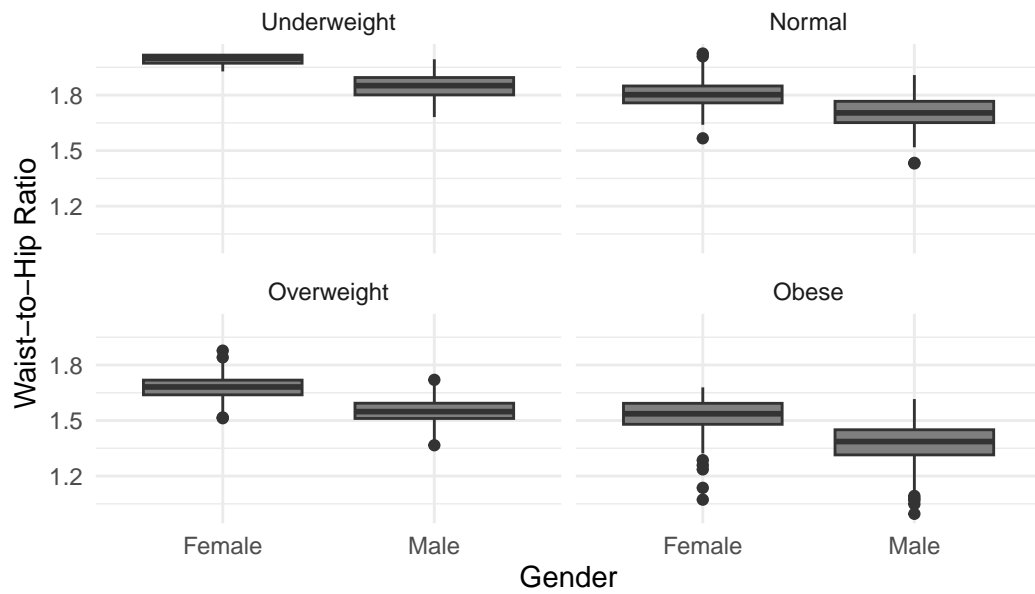
The following table shows the relationship between the new `fat_percentage_category` against existing the `bmi_category`.
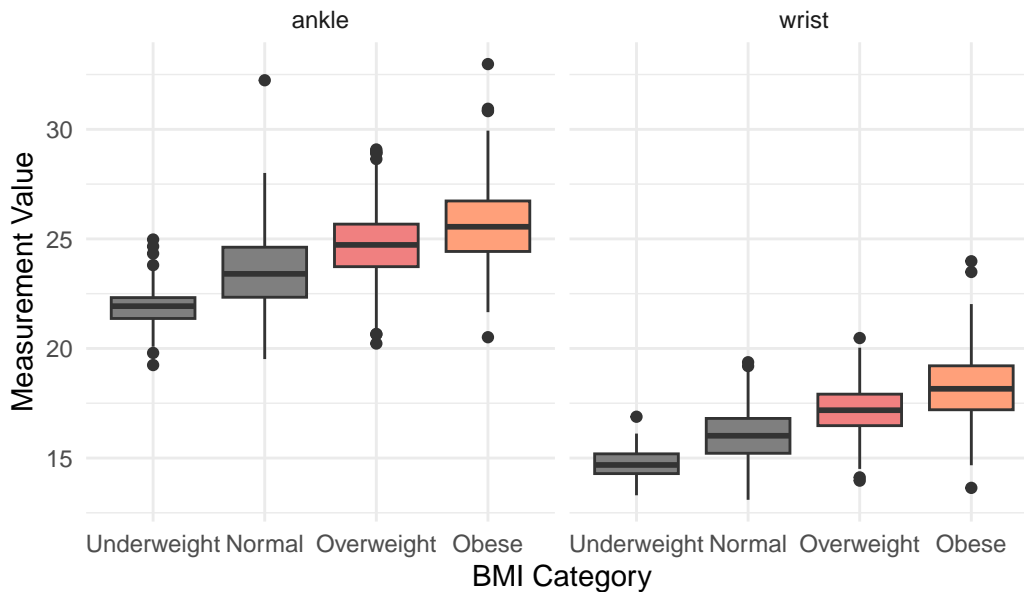

Proportions of BMI Categories by Gender
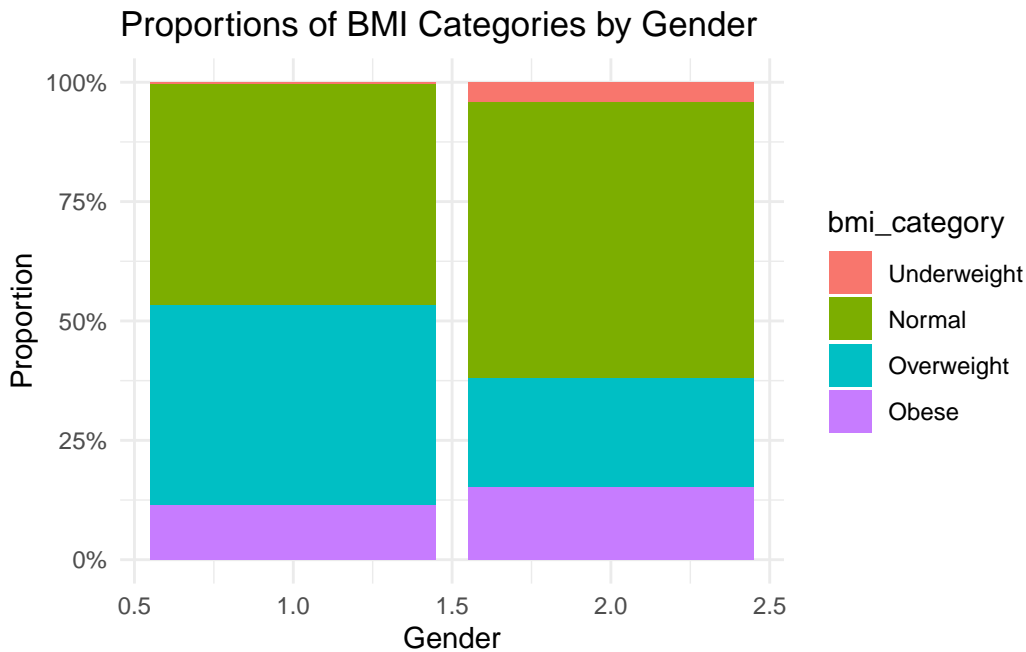
Distribution of Waist–to–Hip Ratio by Gender

Comparison of Ankle and Wrist Measurements for Overweight a
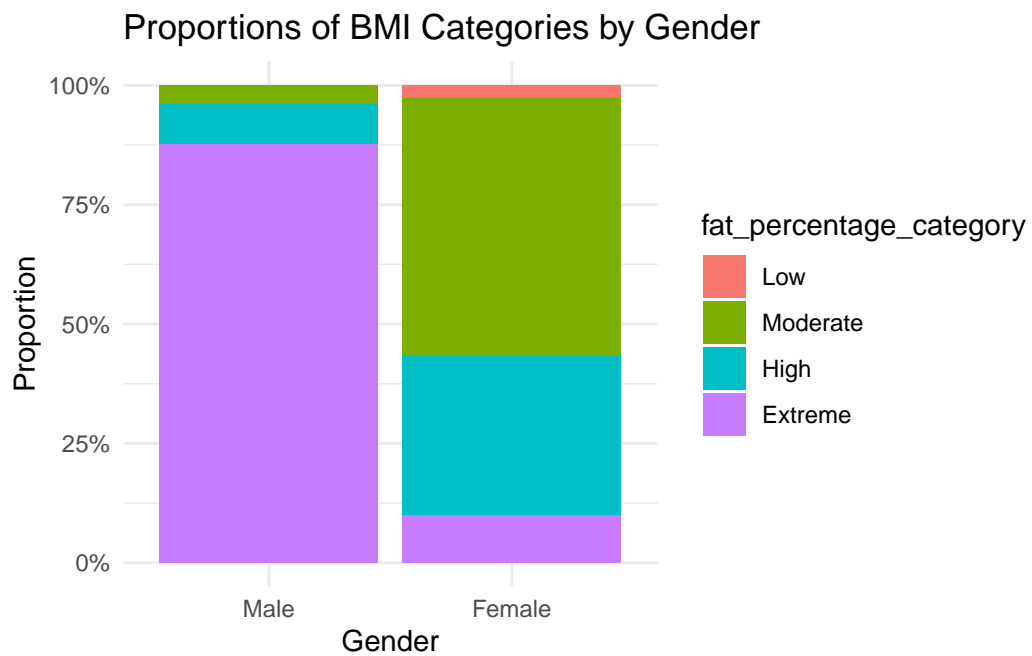
2.1.3 Measurement of Data

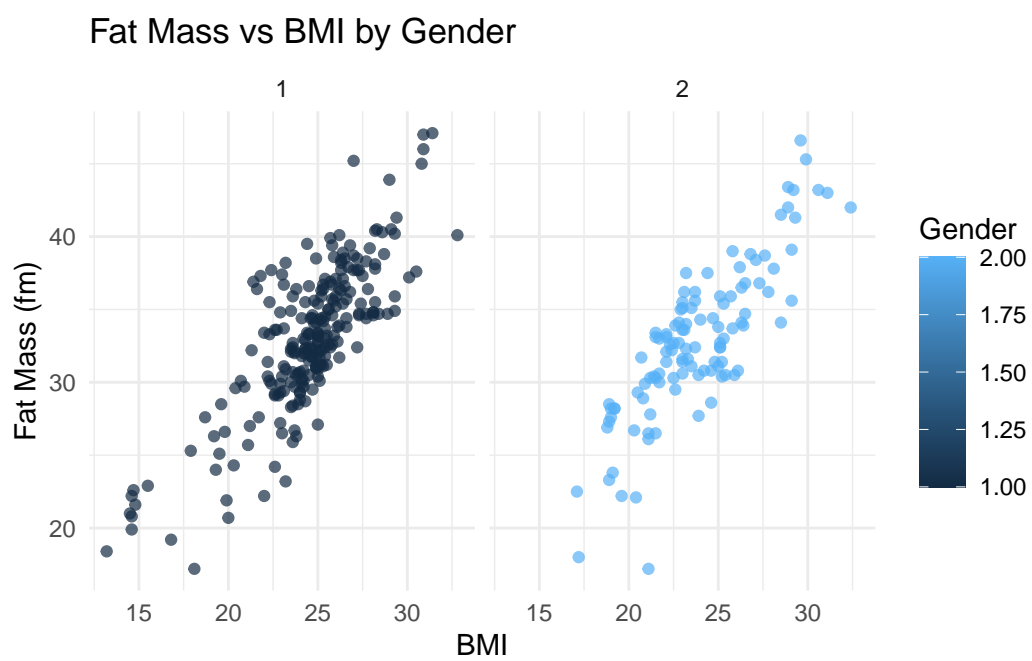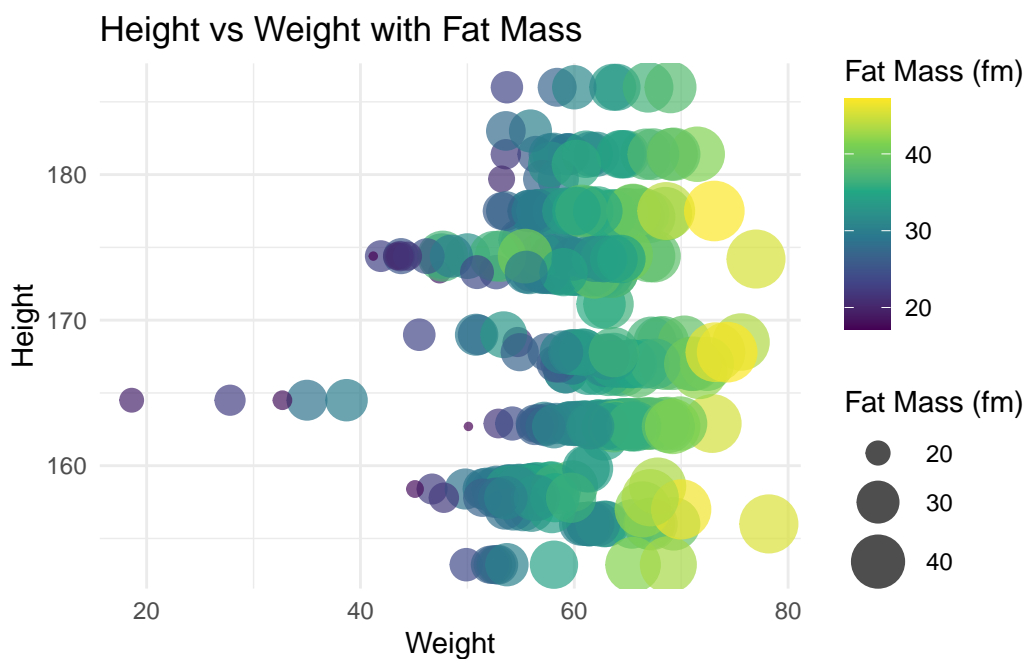## 2.2 Body Composition Data - Isaac Kuzmar Et Al.

The data set compiles the body measurements of subjects aged, 18 and 60, and specifically with a desire to lose weight and improve body image. The participants resided in Barranquilla, Colombia and consisted of 234 males and 111 females. Medical exclusions were made while recruiting the subjects, such as pregnant women or people with medical pacemakers. The bodily measurements such as fat mass in kilograms and fat free mas in kilograms were determined using the Tanita MC-780 [reference] body composition analyzer.

2.2.1 The variables of my use:

The main variables that I considered fat mass percentage fm. This variable is obtained by dividing the fat mass in kilogram by the total weight in kilograms. The fat mass percentage is considered the main predictor for fat mass categories. As the body composition is more accuarate at measuring the actual amount of excess fat a person carries and can differentiate it from other weights such as fat free muscle mass [reference the variable] #### graph what i will be using in my model

2.2.2 The estimand for the model

Proportions of BMI Categories by Gender

Height vs Weight with Fat Mass

Height vs Weight with Fat Mass

Fat Mass vs BMI by Gender

Smoothed Trends of FM and BMI by Age
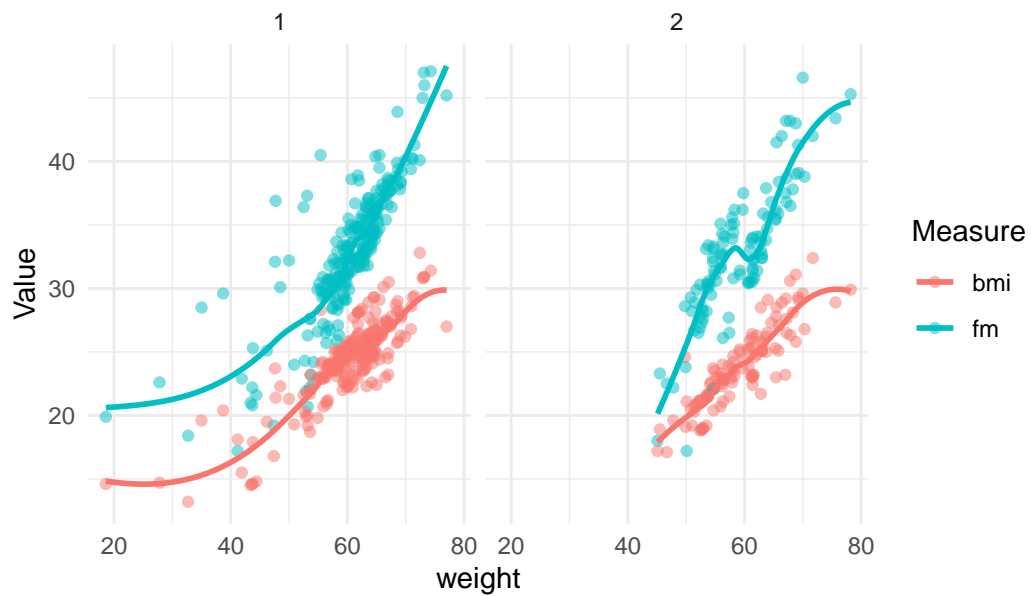
- talk about the crowding for men

Comparison of Fat Percentage Categories Across Data1 and

Proportional Comparison of Fat Percentage Categories Across

Overview text

2.2.3 Measurement of Data

3 Models

Both models, the Measurements Model and the Body Composition Model have the same independent variable that is based on different proxies. For a gross estimation, the primary proxy for each model against Fat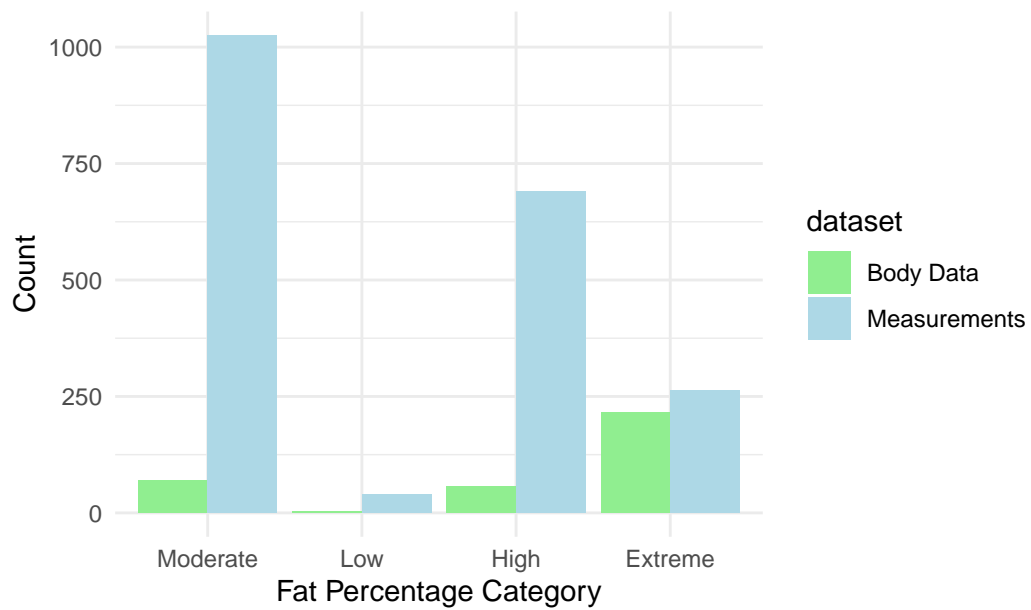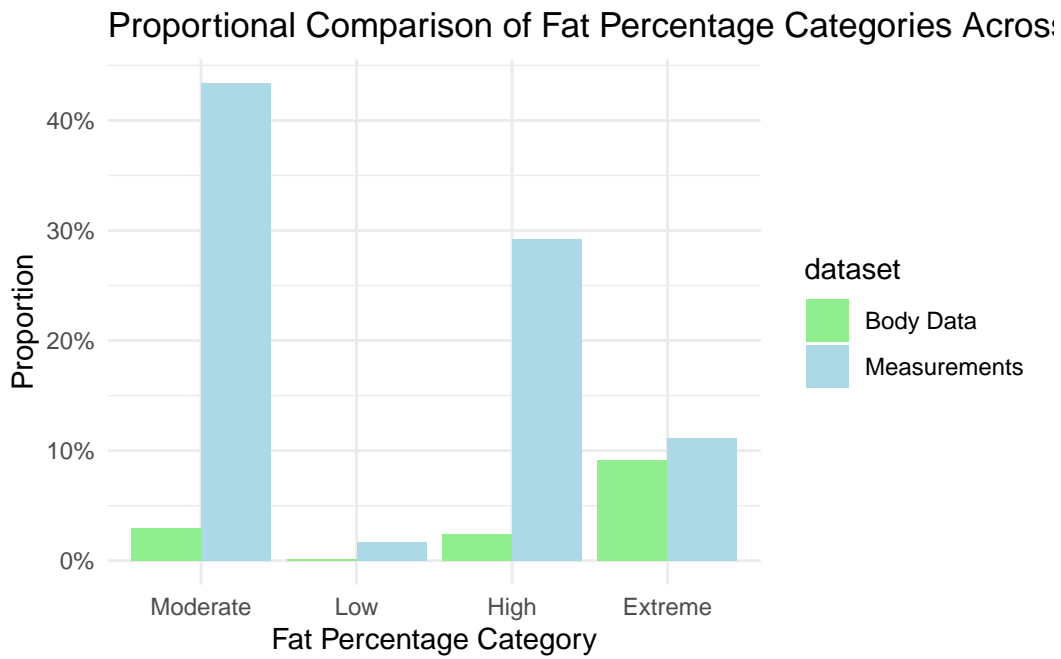 Percentage Category is included in Table 2 below. Reasons behind this decision are discussed in greater detail in each model section. Both models are Multilevel Regression Models to predict `fat_percentage_category` into categories of "Low", "Moderate", "High" and "Extreme" levels. I used the `nnet` (Ripley (2023)) package in R (R Core Team (2023)) to create both multilevel regression models. Model diagnostics including error checks is included in Appendix Section B along with more elaborate explanations, tables and figures.

Table 2: Fat Percentage Category against proxies: BMI Category and Actual Fat Percentage

Fat Percentage Category	BMI	Fat Percentage (Male)	Fat Percentage (Female)
Low	Underweight	<8%	<21%
Moderate	Normal	8-20%	21-33%
High	Overweight	21-25%	34-39%
Extreme	Obese	>25%	>39%

3.1 Measurements Model

I aimed to provide a categorization of body fat levels in an individual based on simple measurements. By using measurements such as height, weight, waist-hip-ratio (WHR), and ankle and wrist circumferences, I sought to capture a person's body composition as accurately as possible. I expected a positive relationship between excess body fat levels and WHR. Additionally, height is an important factor to consider in capturing fat distribution — specifically, a shorter person with a high WHR would be at a higher risk of obesity-related diseases. As men are more likely to accumulate fat in the abdominal region, while women tend to store fat in the hip area. To capture natural body shapes and contours, the height-hip-ratio is also a dominant predictor in the model.

In this model, the dependent variable is `fat_percentage_category`. The model is set up as follows

$$\log\left(\frac{P(Y = k \mid X)}{P(Y = \text{Normal} \mid X)}\right) = \beta_{0k} + \beta_{1k} \cdot \text{height} + \beta_{2k} \cdot \text{gender} + \beta_{3k} \cdot \text{waist\_hip\_ratio} + \beta_{4k} \cdot \text{height\_hip\_ratio}$$

$$(1)$$

Where

- $P(Y = k \mid X)$ is the probability of the response variable Y being in category k (where (k 1, 2, …, K)) given the predictors X.
- Normal is the reference category
- β_{0k}, β_{1k}, β_{2k}, β_{3k}, and β_{4k} are the coefficients corresponding to the intercept and the predictors for category (k).
- The predictors in this model are `height`, `gender`, `waist_hip_ratio`, and `height_hip_ratio`.

3.1.1 Model Justification

It is incredibly challenging to differentiate the body composition of a person based solely on their total weight and height only. However, to increase the accuracy of the classification, this model uses the `waist_to_hip` ratio, which is especially an important indicator for fat accumulation across both genders. The model also uses `height_hip_ratio` to account for different body types. While height is an important physical attribute, simply using the height variable does not effectively capture how a person's body is composed horizontally. So, I chose to employ the `height_hip_ratio` ratio to provide more classification between body types. Additional features such and ankle and wrist are used to further differentiate between the categories. As this model uses BMI as a proxy for body fat, using both height and weight as interacting predictors produced no difference in the new predicted categories of `fat_percentage_category` as the same variables were indirectly used as proxies to assign `fat_percentage_category` using the original BMI formula. Therefore, I chose to only consider height. Another reason for opting out of weight as a predictor, was again, the inability of the model to differentiate between fat and muscle mass. So removing weight altogether was successful in the new classification of fat level.

3.1.2 Assumptions and limitations

Assumptions:

his model assumes that there is a positive relationship with the higher classification of the fat percentage category and the fat storing tendencies in males and females. Other physical ailments that would cause an increase in the circumference measurements without increasing the overall fat in the body is not considered. This model assumes that females with a higher WHR and males with a higher height-hip ratio are classified higher in the fat categories.

Limitations:

The training data set for this model does not include the age of the participants. This is a huge drawback when it comes to assigning risk labels as aging changes how fat is perceived in the bod. Hence, this model would not be able to differentiate between age groups and would work poorly if this information is included. I also considered estimating actual the fat percentage instead of fat percentage *category*. However, to increase the comparability across both models, I chose to use a categorical variable instead.

3.2 Body Composition Model

The model fits a multinomial logistic regression model to predict the fat percentage category (`fat_percentage_category`). The model is defined as follows under the assumption that "Normal" is the baseline category of fat classification:

$$\log \left(\frac{P(Y = k \mid X)}{P(Y = \text{Normal} \mid X)} \right) = \beta_{1k} \cdot \text{age} + \beta_{2k} \cdot \text{height} + \beta_{3k} \cdot \text{gender} + \beta_{4k} \cdot \text{fm} \qquad (2)$$

Where:

- $P(Y = k \mid X)$ is the probability of the response variable Y being in category k (where ($k \in \{$Low, High, Extreme) given the predictors X.
- Normal is the baseline category (the omitted category), meaning the model compares all other categories to "Normal".
- β_{1k}, β_{2k}, β_{3k}, and β_{4k} are the coefficients corresponding to the predictors for category (k), relative to the baseline category "Normal".
- The predictors in this model are `age`, `height`, `gender`, and `fm` (which stands for fat mass percentage).

This model allows us to assess the relationship between the categorical outcome (`fat_percentage_category`) and the continuous and categorical predictors (`age`, `height`, `gender`, `fm`), estimating the probabilities for each category of fat percentage based on the input predictors.

3.2.1 Model Justification

The equation for the model expresses the **log-odds** of being in a given category of `fat_percentage_category` with respect to the baseline category "Normal". For each non-baseline category, the model estimates a separate set of coefficients that describe the log-odds of being in that category compared to being in the "Normal" category. The existence of excess fat in a person has been associated with increased obesity related risks. This model primarily focuses on actual fat in a person and is the main predictor variable of interest. I have considered only two other predictors: `gender` and `height` of a person as there is physiological difference in how access fat affects males and females of persons of different height across different ages. So, `age` is also considered as a predictor variable. Additionally, there are different thresholds to classify a person based on the same fat percentage. As seen in the previous model, I want to study the precise difference in estimation of fat levels when weight is not considered holistically. Hence, `gender` and `height` were considered to be comparable with the Body Measurements Models. I had also considered using muscle mass of the subjects. However, the data was skewed toward obese people as this data sets was

specifically consisted of people who wanted to improve body image. I decided to omit muscle mass as it interacted with the fat mass variable and could be indirectly interpreted as total body weight, suggesting over fitting.

Similar to the Body Measurements Model, this model also considers `fat_percentage_category` to be the estimand. According to the study [], the two sexes have different thresholds for different ages, suggesting linkages to different metabolic needs based on different phases of life. Hence, `fat_percentage_category` is initially classified using `gender`, `age` and `fm`.

3.2.2 Assumptions and limitations

Assumptions:

This model assumes that there is an obvious difference in body measurements and fat storing capacities in males and females, hence gender is an important predictor and also affects the proxy values of fat percentage owing to the calculation based on WHR.

Limitations:

This model is also built on data that was generated by the ABS simulator so it may be inherently working with an error. The error that it could cause in this paper is not accounted for and is considered the true measurement of the test subject.

4 Results

Our results are summarized in Table 3 and Table 4. Our model's results and interpretation:

Table 3: Summary of Measurements Model

	(Intercept)	height	gender	waist_hip_ratio	height_hip_ratio
Low	-34.51538	-0.1646760	3.023376	-19.16317	38.06556
High	25.75761	0.1434481	-2.105247	16.86766	-36.92874
Extreme	43.88456	0.2936878	-3.265845	36.32437	-75.99440

From Table 3 we can see that the intercept for the Low category is negative. Without considering any additional predictors, the model has the highest probability of classifying a person in the Extreme category as its interest is high and positive. As all classifications are considered against the Normal category, the model is most likely to classify a person as having Extreme fat levels and very unlikely to classify them as Low. This data set has poor representation of underweight people and as it is composed of people with a desire to improve body image, this makes sense, as the people with higher fat percentages are more likely to participate in such a study. `waist_hip_ratio` has a negative correlation with the levels of the fat percentage category prediction. High `waist_hip_ratio`s are associated with higher body fat percentage. So the lower a person's `waist_hip_ratio` is, the less likely the model classifies them in the `high` category. On the other hand, `height_hip_ratio` has a positive relation with the classification levels. `height_hip_ratio` takes into a consideration the overall mass (and fat) distribution of a person vertically. So if a person has a more evenly distributed composition, they would have a higher `height_hip_ratio` ratio with a higher chance of being classified in a lower fat category. Despite the model having very few subjects that were underweight, the classification between Underweight and Normal BMI values has improved with more diverse factors. It is likely that Normal BMI categories are also being correctly identified with lower fat levels. There is a negative correlation with the heights of the subjects suggesting that taller people are less likely to be classified with higher levels of fat. Detailed interpretation of this fact is later discussed in the Results Section.

Table 4: Summary of Body Mass Model

	(Intercept)	age	height	gender	fm
Low	17.48090	1.8892595	-0.1293502	69.18352	-9.494383
High	-53.10359	-0.2158061	0.1345077	-43.06833	3.728789
Extreme	-110.98075	-0.6023056	0.3064053	-83.05931	6.797614

For the low category relative to Normal, age and height did not have a significant influence despite initial speculation. Both variables have the β coefficients of $\beta = 1.889$ and $\beta = -0.129$

respectively. However, gender had significant influence on the log-odds of being classified into this category. As males and females are regarded as 1 and 2 in the binary category, `gender`, the $\beta = 69.183$ coefficient indicates that males are more likely to fall into the low category. Naturally, there exists a negative coefficient between fat mass percentage with $\beta = -9.494$

For the High and Extreme categories, age and height had weaker influences. However, gender has a lower log-odd ($\beta = -43.068$ and $\beta = -83.059$ respectively) .Suggesting a negative association between the binary `gender` variable. Hence, females have a higher influence of being classified into these categories over males.

These model results highlight the nuanced relationships between the adipose and anthropocentric measurements. For the Body Composition Model, `gender` and `fm` exhibit the highest influence.

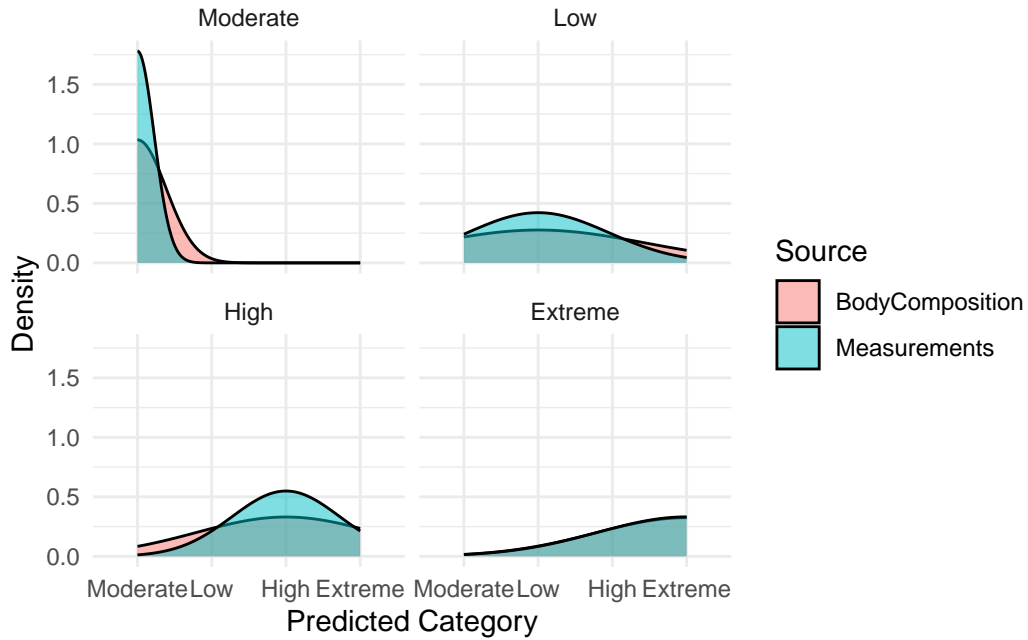## 4.1 Predictions based body circumferences overgeneralize the population

Figure 2: Density Plot of Predicted Categories by Source

From the prediction density chart by each model per fat percentage category, we see that the Measurements model has higher peaks at the each classification. This suggests that more people are being classified into a similar category only based on their physical traits such as height and Waist-to-Hip ratio. The peaks of each category from the Body Composition is accurate in identifying the major classification. However, the distribution not being as concentrated as from the Measurements model suggest that the diversification of different body types is being identified properly.
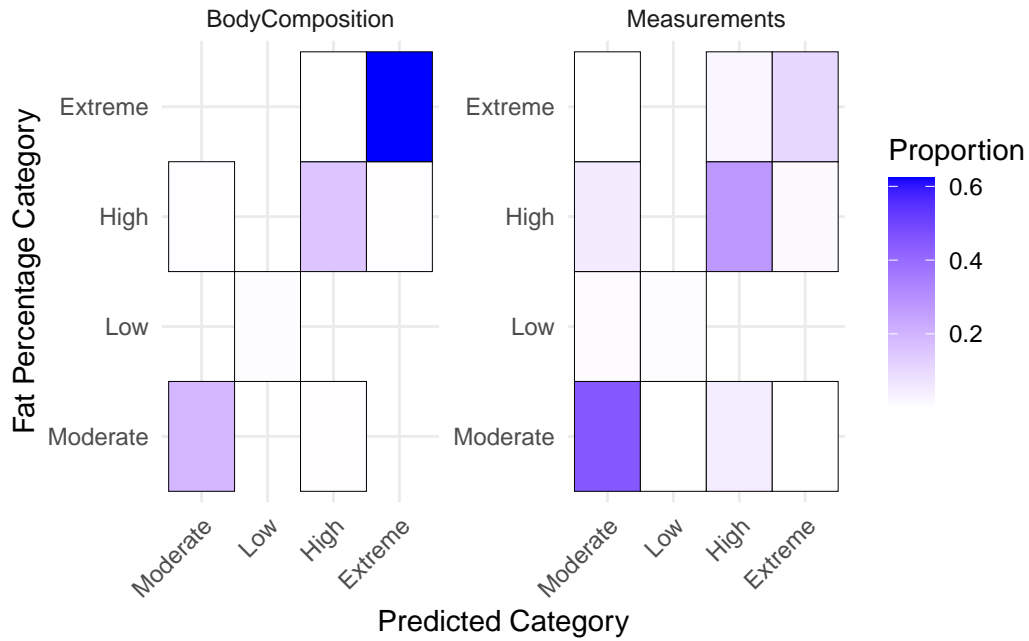
18

Figure 3

The Confusion Matrix shows the difference in the prediction of the fat percentage category by each model. Here, we can see that the Body Composition Data is mostly consistent with its prediction apart from a few key nuances. The diagonal values in the Body Compositions show high correlation between the body fat percentage and the classification. This is expected as this model serves as a comparison to the Measurements Model. However, it can be noted that there is a spill between Moderate and High levels of body fat. The model predicts the Extreme category the closest. On the other hand, there is more obvious spill over between categories using the measurements model. The diagonal values have lighter colours based on the Confusion Matrix Colour Legend. This could suggest that the introduction of adipose and anthropocentric differs in predicting fat in the body as compared to its proxy, BMI. The same hate map faceted by gender, highlights that the BMI proxy in Measurements Model seems to be under performing for females causing more classifications. This could be alluded from the fact that BMI was created with only male subjects under consideration [todo reference].

4.2 What BMI fails to capture

From Figure 5, we see that for a wider range in weight, there is a higher proportion of people who are classifed in the same BMI category. From the Body Composition Model, these categories like on the lower range of the BMI weight thresholds. Increase in fat weight moves people between Fat Classification but they remain in the same BMI category for a longer range of their weight. This shows the lack of factors involved in the BMI classification. Additionally,
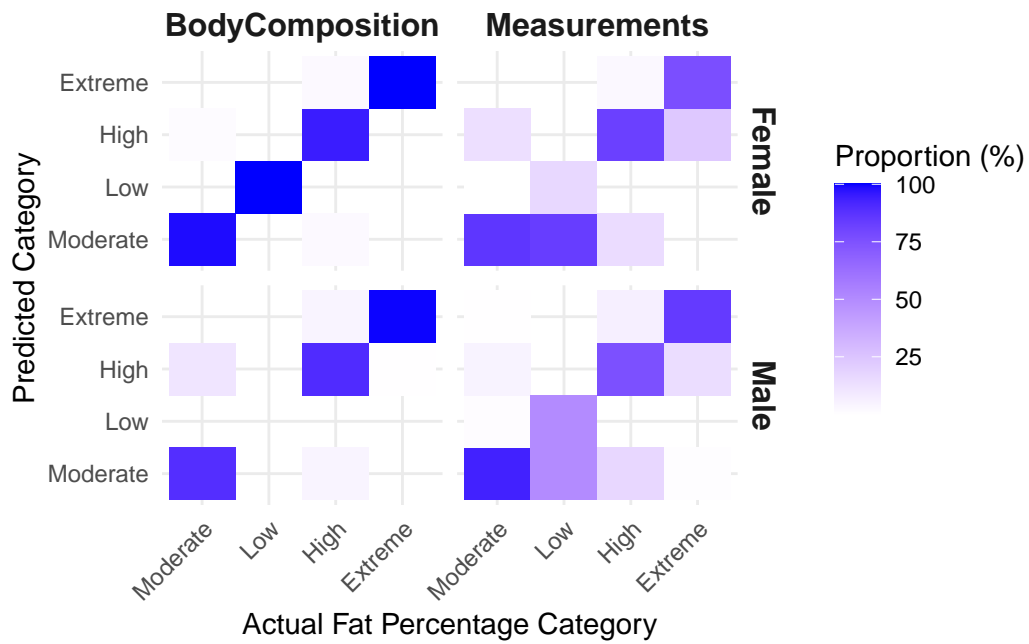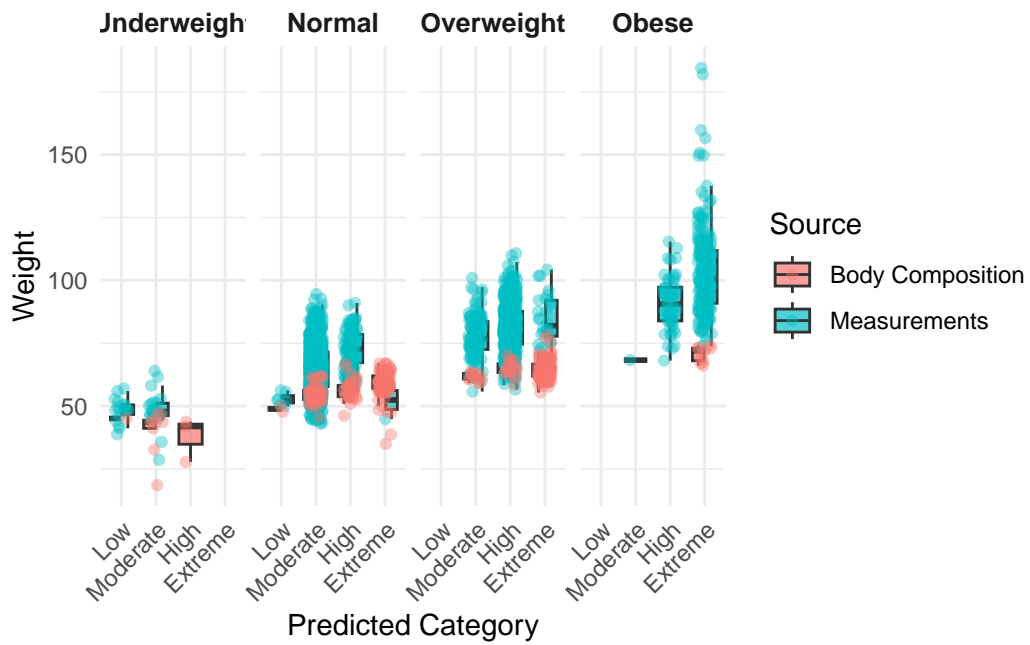
19

Figure 4

Figure 5

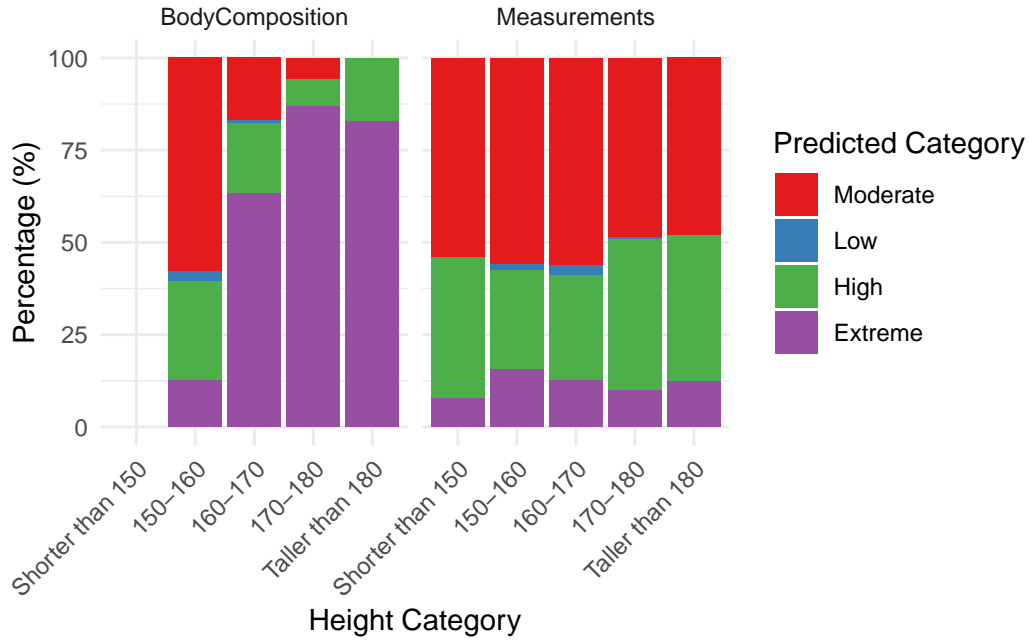## 4.3 Body Fat classifcation for different heights classes

Figure 6

Building on the previous point of BMI failing to correctly capture physical attributes of body composition, this graph emphasises how the Moderate Category is persistently overclassiffied across height groups. In the Measurements Model graph, all height groups share similar proportions of Fat Category classifications. More than 50% of the people below 170CM and slightly less than 50% people taller than 170CM are classified as Normal Fat Category. However, this overclassication is debunked by the Body Composition Model Graph as almost all of the height classes have a major classification in the Normal category. In fact, height classes above 160CM have a collective proportion of greater than 75% for High and Extreme fat levels. As the Body Composition Model specifically disregards weight holistically and considers pure fat only, we can see that tall people have an increased chance of being classified as Normal fat category. Subsequent health risk identifications are delayed for these groups of people and is later discussed in greater detail [here todo.] Additionally, following the original BMI formula, taller people are at advantage as their weight is spread further along their height. However, the difference in proportions of the classifications across both models highlights this specific downfall of the BMI formula.

4.4 Difference in Gender Performance

Figure 7 shows the performance of each of the models against the `gender` predictor. For females, the proporitons for all categories are higher in the Body Composition Model suggesting that the
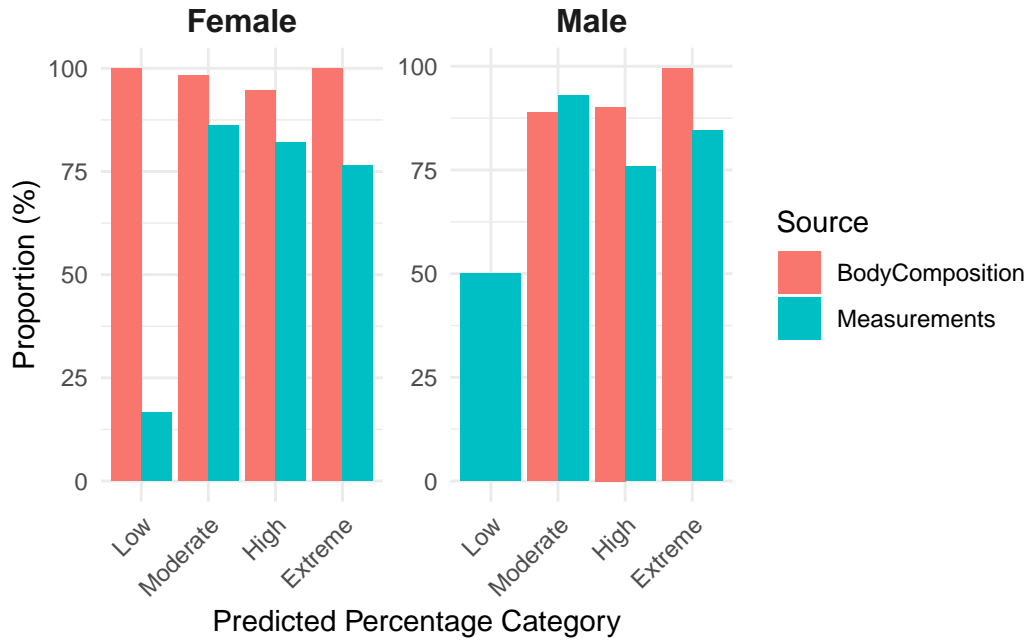
Figure 7

model is able to correctly classify the fat percentage category based on the predictors `gender` and `fm`. However, the Measurements Model has lower classifications suggesting that there is different classification based on the BMI proxy and the predictors `gender`, `waist_hip_ratio` and `height_hip_ratio`.

5 Discussion

5.1 Potential Health Risk Identification

This paper's main goal was to compare how accurate are body measurements to estimate fat percentage at home. Other ways of estimating a person's body fat include scans like DEXA (Kuzmar2020). These scans are expensive and inaccessible. Most people turn to measuring BMI's to estimate their general level of health. The Government of Canada (Canada (2024)) also advices that adults use BMI to estimate their health. As seen in this paper, this inherently flawed as BMI fails to capture the composition of a body. However, even the introduction of simple measurement like Waist-to-Hip ratio, shows that there is a general improvement in differentiating fat levels. The Waist-to-Hip ratio is especially important between the two sexes as fat accumulation pattern differs, which is not even considered in the BMI formula. Additionally, as a higher proportion of people are classified as Normal using BMI, without

22

failing to recognise that having a high level of fat despite having a lower BMI is still danger-ous as it could potentially prevent them from seeking early care against illnesses like insulin resistance. By popularizing a more accurate system, we could urge people to take earlier pre-ventive care. This opens up the path for the need for more sophisticated medical interventions. According to Adab, Pallan, and Whincup (2018), ratios such WHR are beneficial to identify cardiovascular health risks due to adipose tissue fat accumulation. However, due to the lack of standardization in current medical care, this ratio is yet to gain popularity over the BMI formula.

It also maybe understood that WHR shows a better performance for certain demographics only. In the study done by Li et al. (2006), the WHR showed no increased association with cardiovascular diseases in Overweight and Obese men. However, for other demographics such as women, there is parallel increase in the risk of cardiovascular diseases with increase in WHR given the same BMI category. Nearly a double association with cardiovascular diseases was seen in Normal and Overweight women who had a higher WHR than women who had a lower WHR. This relation could applied to Normal BMI only for men. As the WHR is non-standard and of value to a smaller demographic of the general population, a wide clinical acceptance has not been seen for it.

5.2 What should athletes use?

BMI is a crude way of classifying people. Athletes are especially at risk at being mis-classified despite a lower body mass weight but a higher muscle mass. According to [], Athletes have

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Weaknesses and next steps

- the BodyM dataset does not include age. An older person has lower metabolism [] and a higher chance of collecting fat on their abdomen.

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Models Validation

B.1 Posterior predictive check

we compare the posterior with the prior. This shows...

References

Adab, Peymane, Miranda Pallan, and Peter H Whincup. 2018. "Is BMI the Best Measure of Obesity?" *BMJ* 360. https://doi.org/10.1136/bmj.k1274.

Amazon Web Services. 2023. "AWS Bodym." https://registry.opendata.aws/bodym/.

Canada, Health. 2024. "Body Mass Index Nomogram." https://www.canada.ca/en/health-canada/services/food-nutrition/healthy-eating/healthy-weights/canadian-guidelines-body-weight-classification-adults/body-mass-index-nomogram.html.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://github.com/sfirke/janitor.

Kuzmar, Isaac, and Zalabata. 2020. "DATASET FOR NEW BODY FAT MEASUREMENT." https://doi.org/10.6084/m9.figshare.12982223.v2.

Li, C., G. Engström, B. Hedblad, S. Calling, G. Berglund, and L. Janzon. 2006. "Sex Differences in the Relationships Between BMI, WHR and Incidence of Cardiovascular Disease: A Population-Based Cohort Study." *International Journal of Obesity* 30 (12): 1775–81. https://doi.org/10.1038/sj.ijo.0803339.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'.* https://CRAN.R-project.org/package=arrow.

Ripley, B. 2023. *Nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models.* https://CRAN.R-project.org/package=nnet.

Ruiz, Nataniel, Miriam Bellver Bueno, Timo Bolkart, Ambuj Arora, Ming Lin, Javier Romero, and Raja Bala. 2022. "Human Body Measurement Estimation with Adversarial Augmentation." In *International Conference on 3D Vision 2022.* https://www.amazon.science/publications/human-body-measurement-estimation-with-adversarial-augmentation.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.