Example Manuscript Template for a Data Analysis Project

Doreen Kibuule Kalembe

Warning: package 'here' was built under R version 4.4.2

**Authors**

* First Autor (ORCID: 0000-0000-1234-5678)
* Second Author

**Author affiliations**

1. College of Public Health, University of Georgia, Athens, GA, USA.

These authors contributed equally to this work.

Corresponding author: kelvinbwambale8@gmail.com

Disclaimer: The opinions expressed in this article are the author’s own and don’t reflect their employer.

# 1. Summary/Abstract

TITLE: Sex-based analysis/Differences in TB Immune Activation.

Tuberculosis is one of the leading causes of death in the world.According to NIH reports, tuberculosis (TB), afflicts more human males than females, with a ratio of 1.7 males to female. The disease is still endemic in most of the sub-Saharan African countries. In Uganda, for every one case of tuberculosis reported, there are at least three cases of tuberculosis that are not identified and its because of these cases that the disease remains prevalent in the country. Previous studies show that there are differences in the immune activation of TB among men and women. In this study, we aim to explore possible underlying hormonal and genetic mechanisms that might contribute to differences in the immune activation among females and males.

#Methods We utilized data from a tuberculosis public repository available through Tidyverse (https://dataverse.unc.edu/dataset.xhtml?persistentId=doi:10.15139/S3/AYOFEU). The dataset comprises 60 individuals - 30 males and 30 females. All participants were HIV-negative and tuberculosis-free at the time of enrollment.

The key variables in our analysis include:

CD4 immune activation count Disease severity indicators (extent of lung disease and number of lungs involved) Fat in kilograms Sex/Gender BMI Age of participant These demographic and clinical variables form the foundation of our investigation into tuberculosis-related immune responses.

#Proposed Analysis Method.

The annalysis was conducted in R version 4.4.2. This was done to Compare immune activation levels between males and females.

We computed the male-to-female differences in disease severity using cd4 immune activation counts.We correlated immune markers with disease severity measures/Correlation between activation levels and disease severity. We did this using linear regression and machine learning models to predict which sex will show the most results. We also adjusted for potential confounders like age.

#Results # Results

Our analysis revealed a statistically significant difference in CD4 immune activation counts between males and females. On average, females showed lower CD4 activation counts than males (mean difference = 413) indicating less immune activation in females and a true difference in the body fat between females and males (mean difference = 17.76 95%CI (7.097298 29.972199).

Linear regression models adjusted for age, BMI, and fat mass demonstrated that male sex was independently associated with higher CD4 immune activation (β = 692.58 95% CI = [-786.97, 2172.13]).

Disease severity indicators were more strongly associated with immune activation in males than in females, suggesting a sex-based difference in how immune response relates to disease burden.

The LASSO and Random Forest models identified BMI, sex, CD4+, LBM as the most important predictors of immune activation. These findings support the hypothesis that sex-based physiological differences, particularly in fat composition, contribute to variability in immune activation and disease severity in tuberculosis.

#conclusions. Sex differences in TB is contributed by more than just hormonal influence. Our study shows that TB is more severe in men than it is in women. Regardless of the medical seeking behaviors of women, women tend to a have a high body fat percentage more than men.The higher the fat percentage an individual has, the lower the risk of progression of the tuberculosis infection to disease hence lowering mortality rates among females compared to men. In addition, women tend to have more lean body mass compared to men hence having a lower chance of suffering from adverse effects of TB compared to men.

# 2. Introduction

## 2.1 General Background Information

Tuberculosis (TB) remains a significant global public health challenge. According to the World Health Organization (WHO), reports show that 10 million new cases and 1.4 million deaths in 2023 (1) were as a result of Tuberculosis. The spread of M. tuberculosis, as the pathogen of TB, has long been hypothesized to occur more often in the household than in the community(2). While most of the M.Tuberculosis transmission in the community is attributed to the unknown contact networks that occurs between individuals, this accounts upto 90% of the transmission. The recent outbreak of Tuberculosis in the united states clearly explains the contact network phenomena. Overall, TB accounts for one quarter of the annual deaths that happen worldwide. An estimated 10.0 million people fell ill with TB in 2018. Geographically, Southeast Asia ranked highest (44%), followed by Africa (24%) and the Western Pacific (18%), with lesser percentages in the Eastern Mediterranean (8%), the Americas (3%), and Europe (3%). Among women of reproductive years, TB is the largest cause of death (3). Global Strategies made by WHO, aim at ending the TB epidemic, with targets to reduce the number of deaths caused by TB by 95% and to decrease new cases by 90% from 2015 to 2035(4).

Mycobacterium tuberculosis infection primarily spreads when a person shares airspace with someone who has active tuberculosis and inhales infectious droplet aerosols. Since the bacterium transmits through respiratory routes to susceptible individuals, the household of an infected person becomes a critical environment for intense M. tuberculosis transmission, where most new infections and cases can be detected and treated. Research reports indicate significant differences in tuberculosis responses between men and women. Notably, these studies suggest that women face greater barriers to early detection and treatment of tuberculosis compared to men, potentially impacting disease outcomes and progression.

The World Health Organization, among other institutions, attributes differences in tuberculosis trends between men and women to several factors: access to care and treatment, pregnancy effects, body fat mass, and gender roles that diminish women’s social capital, thereby lowering incidence among them (5). Worldwide, epidemiological data indicate that tuberculosis (TB) rates among HIV-negative men have consistently been higher than among women for many decades (6). While biological mechanisms likely influence these differences, previous studies demonstrate that men experience higher incidence and prevalence of disease, potentially leading to poorer treatment outcomes compared to women. According to an analysis conducted in 2000, TB prevalence among males exceeded that among females in 27 (93%) of 29 prevalence surveys conducted across 14 countries between 1953 and 1997 (6).

Other studies have highlighted that body fat plays an important role in supporting immune function, with women typically maintaining higher fat storage than men. This increased fat storage may contribute to women’s stronger immune responses. While these differences are partly attributed to women’s better healthcare-seeking behaviors, factors such as fat distribution, hormonal profiles, body composition, lean tissue mass, and other biological characteristics differ significantly between sexes, particularly at disease presentation. The objective of this study is to determine whether higher immune activation counts correlate with greater infection severity or illness progression.

## 2.2 Description of data and data source

The data we used for this project was got from the TRAC dataverse and is fully accessible to everyone. This is considered a safe place where the the Tuberculosis data from the TRAc project is kept and can be used by anyone who wants contribute to TB research. More details about the data can be found [here](https://dataverse.unc.edu/dataset.xhtml?persistentId=doi:10.15139/S3/AYOFEU). The data is publicly available and can be accessed by anyone.

## 2.3 Questions/Hypotheses to be addressed

Since the balance of fat and lean tissue differ in men and women especially at the time of tuberculosis presentation, the objective of this study will be to find out whether there are differences in the Tuberculosis immune activation between males and females. We shall hypothesize on the more immune activation there is, the more infected / sicker the person.

# 3. Methods

Study Design

This study utilized a cross sectional study design analyzing data from 2017.Sixty TB-confirmed participants (30 males, 30 females, aged 15+) were recruited from Mulago Hospital.

###inclusion and exclusion critera

Our study included patients aged 15 years and older experiencing their first episode of tuberculosis. To maintain data integrity, we excluded individuals receiving hormonal therapy or using hormonal contraceptives. Additionally, patients with comorbidities including diabetes mellitus, cancer, or HIV were excluded. Our strict exclusion criteria ensured a focused research population by omitting all individuals under 15 years of age, those with recurrent TB episodes, patients taking any form of hormonal therapy or contraception, and those diagnosed with diabetes mellitus, cancer, or HIV.

The variables of interest were extracted from the TRAC dataset and exported to an MS Excel sheet. The outcome variable was CD4+ immune activation counts. The primary independent variable was sex (cis-gender) and this was categorized as male and female. The covariates were age at diagnosis,BMI, Fat in Kg, lean body mass in kg(LBM in kg), CD4+ and CD4 immune activation counts. other variables like fat mass index (FMI) and fat-free mass index (FFMI) were also obtained.

## 3.1 Schematic of workflow

Chromosomal sex and hormones play a role in the modulation of immune responses to mTB and thereby contribute to the general outcome of disease. The first cells that encounter Mtb are known as Macrophages. These initiate a local inflammatory reaction which may differ between the sexes(7). This causes the long term containment of TB infection to differ between men and women since their adaptive immune responses are different. See (**tbfig-1?**)

|  |
| --- |
| Figure 1: Impact of sex differences on the immune response to Mtb and their influence disease outcome. |

## 3.2 Data import and cleaning

We imported the data from MS Excel to R. We explored the data and found some missing values for some variables. We decided to drop one observation since the person lacked the primary exposure and did not have most of the details we needed for our final analysis. We categorized the participants age into four categories. That is 17-<28, 28-<40, 40-<60 and >=60.

## 3.3 Statistical analysis

The final cleaned data set contains some negative fat values. Negative values occur when the deuterium dose has not had sufficient time to fully equilibrate with body water, or the dose was not completely consumed. This is seen as low deuterium enrichment, resulting in an overestimation of the size of the body water pool and hence high FFM (FFMI = LBM / height2)and low % body fat.

We applied three supervised machine learning models — linear regression, LASSO regression, and Random Forests — to predict CD4 Immune Activation Count based on Sex, Age, BMI, Lean Body Mass (LBM), Fat mass, and CD4+ cell count. Data preprocessing included median imputation for missing values, dummy encoding of categorical variables, normalization, and removal of highly correlated features. Models were evaluated using RMSE and visualized through observed vs. predicted plots. Hyperparameter tuning was performed using grid search with cross-validation.

The significance of Sex as a predictor of immune activation was assessed in each model. For linear models, coefficients for Sex were examined along with their corresponding p-values. The Random Forest model’s variable importance was assessed to identify the predictive contribution of Sex.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

We conducted a thorough data exploration using various visualization techniques and summary tables. Our analysis included box plots, histograms, and correlation plots to identify meaningful patterns in the dataset. We employed scatter plots to examine trends and investigate potential relationships between continuous variables. This comprehensive approach allowed us to uncover several interesting insights within the data.

(**Table1?**) shows the summary statistics of our data variables.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Summary distribution of baseline characteristics among males and females.   |  | F (N=30) | M (N=30) | Overall (N=60) | | --- | --- | --- | --- | | **CD4+** |  |  |  | | Mean (SD) | 394 (217) | 289 (181) | 341 (205) | | Median [Min, Max] | 449 [27.0, 786] | 278 [8.64, 735] | 319 [8.64, 786] | | Missing | 1 (3.3%) | 0 (0%) | 1 (1.7%) | | **CD4 Immune activation count** |  |  |  | | Mean (SD) | 1750 (1340) | 2160 (3170) | 1960 (2430) | | Median [Min, Max] | 1520 [94.0, 7350] | 1210 [84.0, 17900] | 1380 [84.0, 17900] | | Missing | 1 (3.3%) | 0 (0%) | 1 (1.7%) | | **Participant age** |  |  |  | | Mean (SD) | 29.9 (12.8) | 30.2 (6.83) | 30.0 (10.1) | | Median [Min, Max] | 26.0 [17.0, 70.0] | 28.5 [19.0, 48.0] | 28.0 [17.0, 70.0] | | **Fat in kg** |  |  |  | | Mean (SD) | 18.9 (22.4) | 0.348 (21.9) | 9.62 (23.8) | | Median [Min, Max] | 10.9 [-0.192, 81.0] | 4.37 [-96.6, 21.3] | 8.70 [-96.6, 81.0] | | **LBM in kg** |  |  |  | | Mean (SD) | 31.1 (21.0) | 54.9 (21.1) | 43.0 (24.1) | | Median [Min, Max] | 39.6 [-42.0, 45.9] | 50.0 [38.2, 151] | 43.4 [-42.0, 151] | |

Table [Table 2](#tbl-correlation) shows the correlation matrix between immune activation markers and body composition variables.

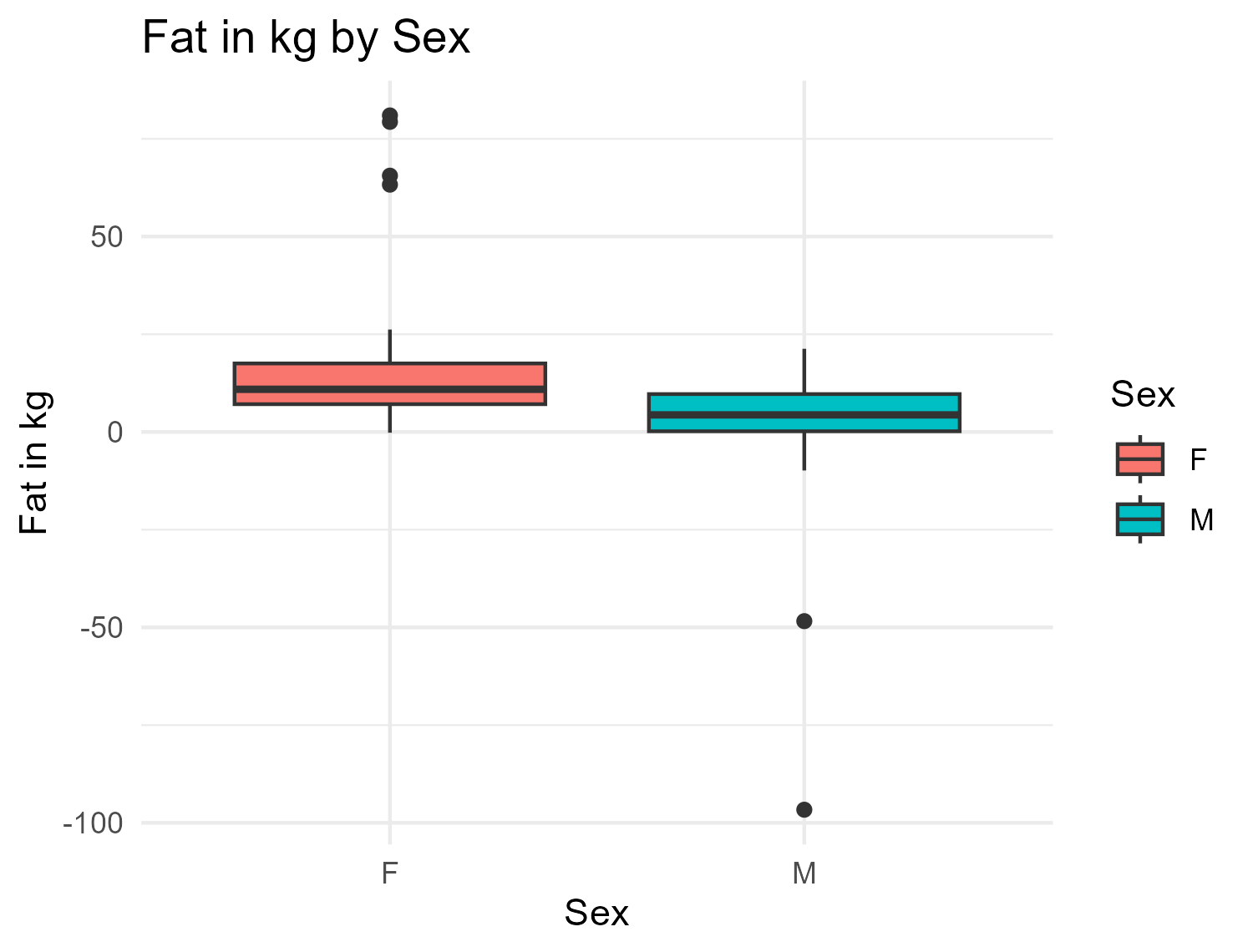
|  |
| --- |
| Table 2: Correlation matrix between immune activation markers and body composition variables.  $`\_data`  Variable CD4+ CD4 Immune activation count Fat in kg 1 CD4+ 1.00 0.22 0.17 2 CD4 Immune activation count 0.22 1.00 0.02 3 Fat in kg 0.17 0.02 1.00 4 LBM in kg -0.18 -0.04 -0.94 5 Participant age 0.08 -0.02 0.11 6 BMI 0.08 -0.15 0.19  LBM in kg Participant age BMI 1 -0.18 0.08 0.08 2 -0.04 -0.02 -0.15 3 -0.94 0.11 0.19 4 1.00 -0.10 0.06 5 -0.10 1.00 0.12 6 0.06 0.12 1.00  $`\_boxhead`  var type column\_label column\_units 1 Variable default Variable <NA> 2 CD4+ default CD4+ <NA> 3 CD4 Immune activation count default CD4 Immune activation count <NA> 4 Fat in kg default Fat in kg <NA> 5 LBM in kg default LBM in kg <NA> 6 Participant age default Participant age <NA> 7 BMI default BMI <NA>  column\_pattern column\_align column\_width hidden\_px 1 <NA> left NULL NULL 2 <NA> right NULL NULL 3 <NA> right NULL NULL 4 <NA> right NULL NULL 5 <NA> right NULL NULL 6 <NA> right NULL NULL 7 <NA> right NULL NULL  $`\_stub\_df`  rownum\_i row\_id group\_id group\_label indent built\_group\_label 1 1 <NA> <NA> NULL <NA> <NA> 2 2 <NA> <NA> NULL <NA> <NA> 3 3 <NA> <NA> NULL <NA> <NA> 4 4 <NA> <NA> NULL <NA> <NA> 5 5 <NA> <NA> NULL <NA> <NA> 6 6 <NA> <NA> NULL <NA> <NA>  $`\_row\_groups` character(0)  $`\_heading` $`\_heading`$title [1] "Table 2: Correlation Matrix of Immune Activation and Body Composition"  $`\_heading`$subtitle NULL  $`\_heading`$preheader NULL   $`\_spanners` [1] vars spanner\_label spanner\_units spanner\_pattern [5] spanner\_id spanner\_level gather built  <0 rows> (or 0-length row.names)  $`\_stubhead` $`\_stubhead`$label NULL   $`\_footnotes` [1] locname grpname colname locnum rownum colnum footnotes [8] placement <0 rows> (or 0-length row.names)  $`\_source\_notes` list()  $`\_formats` list()  $`\_substitutions` list()  $`\_styles` [1] locname grpname colname locnum rownum colnum styles  <0 rows> (or 0-length row.names)  $`\_summary` list()  $`\_options`  parameter scss category type 1 table\_id FALSE table value 2 table\_caption FALSE table value 3 table\_width TRUE table px 4 table\_layout TRUE table value 5 table\_margin\_left TRUE table px 6 table\_margin\_right TRUE table px 7 table\_background\_color TRUE table value 8 table\_additional\_css FALSE table values 9 table\_font\_names FALSE table values 10 table\_font\_size TRUE table px 11 table\_font\_weight TRUE table value 12 table\_font\_style TRUE table value 13 table\_font\_color TRUE table value 14 table\_font\_color\_light TRUE table value 15 table\_border\_top\_include FALSE table logical 16 table\_border\_top\_style TRUE table value 17 table\_border\_top\_width TRUE table px 18 table\_border\_top\_color TRUE table value 19 table\_border\_right\_style TRUE table value 20 table\_border\_right\_width TRUE table px 21 table\_border\_right\_color TRUE table value 22 table\_border\_bottom\_include FALSE table logical 23 table\_border\_bottom\_style TRUE table value 24 table\_border\_bottom\_width TRUE table px 25 table\_border\_bottom\_color TRUE table value 26 table\_border\_left\_style TRUE table value 27 table\_border\_left\_width TRUE table px 28 table\_border\_left\_color TRUE table value 29 heading\_background\_color TRUE heading value 30 heading\_align TRUE heading value 31 heading\_title\_font\_size TRUE heading px 32 heading\_title\_font\_weight TRUE heading value 33 heading\_subtitle\_font\_size TRUE heading px 34 heading\_subtitle\_font\_weight TRUE heading value 35 heading\_padding TRUE heading px 36 heading\_padding\_horizontal TRUE heading px 37 heading\_border\_bottom\_style TRUE heading value 38 heading\_border\_bottom\_width TRUE heading px 39 heading\_border\_bottom\_color TRUE heading value 40 heading\_border\_lr\_style TRUE heading value 41 heading\_border\_lr\_width TRUE heading px 42 heading\_border\_lr\_color TRUE heading value 43 column\_labels\_background\_color TRUE column\_labels value 44 column\_labels\_font\_size TRUE column\_labels px 45 column\_labels\_font\_weight TRUE column\_labels value 46 column\_labels\_text\_transform TRUE column\_labels value 47 column\_labels\_padding TRUE column\_labels px 48 column\_labels\_padding\_horizontal TRUE column\_labels px 49 column\_labels\_vlines\_style TRUE table\_body value 50 column\_labels\_vlines\_width TRUE table\_body px 51 column\_labels\_vlines\_color TRUE table\_body value 52 column\_labels\_border\_top\_style TRUE column\_labels value 53 column\_labels\_border\_top\_width TRUE column\_labels px 54 column\_labels\_border\_top\_color TRUE column\_labels value 55 column\_labels\_border\_bottom\_style TRUE column\_labels value 56 column\_labels\_border\_bottom\_width TRUE column\_labels px 57 column\_labels\_border\_bottom\_color TRUE column\_labels value 58 column\_labels\_border\_lr\_style TRUE column\_labels value 59 column\_labels\_border\_lr\_width TRUE column\_labels px 60 column\_labels\_border\_lr\_color TRUE column\_labels value 61 column\_labels\_hidden FALSE column\_labels logical 62 column\_labels\_units\_pattern FALSE column\_labels value 63 row\_group\_background\_color TRUE row\_group value 64 row\_group\_font\_size TRUE row\_group px 65 row\_group\_font\_weight TRUE row\_group value 66 row\_group\_text\_transform TRUE row\_group value 67 row\_group\_padding TRUE row\_group px 68 row\_group\_padding\_horizontal TRUE row\_group px 69 row\_group\_border\_top\_style TRUE row\_group value 70 row\_group\_border\_top\_width TRUE row\_group px 71 row\_group\_border\_top\_color TRUE row\_group value 72 row\_group\_border\_right\_style TRUE row\_group value 73 row\_group\_border\_right\_width TRUE row\_group px 74 row\_group\_border\_right\_color TRUE row\_group value 75 row\_group\_border\_bottom\_style TRUE row\_group value 76 row\_group\_border\_bottom\_width TRUE row\_group px 77 row\_group\_border\_bottom\_color TRUE row\_group value 78 row\_group\_border\_left\_style TRUE row\_group value 79 row\_group\_border\_left\_width TRUE row\_group px 80 row\_group\_border\_left\_color TRUE row\_group value 81 row\_group\_default\_label FALSE row\_group value 82 row\_group\_as\_column FALSE row\_group logical 83 table\_body\_hlines\_style TRUE table\_body value 84 table\_body\_hlines\_width TRUE table\_body px 85 table\_body\_hlines\_color TRUE table\_body value 86 table\_body\_vlines\_style TRUE table\_body value 87 table\_body\_vlines\_width TRUE table\_body px 88 table\_body\_vlines\_color TRUE table\_body value 89 table\_body\_border\_top\_style TRUE table\_body value 90 table\_body\_border\_top\_width TRUE table\_body px 91 table\_body\_border\_top\_color TRUE table\_body value 92 table\_body\_border\_bottom\_style TRUE table\_body value 93 table\_body\_border\_bottom\_width TRUE table\_body px 94 table\_body\_border\_bottom\_color TRUE table\_body value 95 data\_row\_padding TRUE data\_row px 96 data\_row\_padding\_horizontal TRUE data\_row px 97 stub\_background\_color TRUE stub value 98 stub\_font\_size TRUE stub px 99 stub\_font\_weight TRUE stub value 100 stub\_text\_transform TRUE stub value 101 stub\_border\_style TRUE stub value 102 stub\_border\_width TRUE stub px 103 stub\_border\_color TRUE stub value 104 stub\_indent\_length TRUE stub px 105 stub\_row\_group\_background\_color TRUE stub value 106 stub\_row\_group\_font\_size TRUE stub px 107 stub\_row\_group\_font\_weight TRUE stub value 108 stub\_row\_group\_text\_transform TRUE stub value 109 stub\_row\_group\_border\_style TRUE stub value 110 stub\_row\_group\_border\_width TRUE stub px 111 stub\_row\_group\_border\_color TRUE stub value 112 summary\_row\_padding TRUE summary\_row px 113 summary\_row\_padding\_horizontal TRUE summary\_row px 114 summary\_row\_background\_color TRUE summary\_row value 115 summary\_row\_text\_transform TRUE summary\_row value 116 summary\_row\_border\_style TRUE summary\_row value 117 summary\_row\_border\_width TRUE summary\_row px 118 summary\_row\_border\_color TRUE summary\_row value 119 grand\_summary\_row\_padding TRUE grand\_summary\_row px 120 grand\_summary\_row\_padding\_horizontal TRUE grand\_summary\_row px 121 grand\_summary\_row\_background\_color TRUE grand\_summary\_row value 122 grand\_summary\_row\_text\_transform TRUE grand\_summary\_row value 123 grand\_summary\_row\_border\_style TRUE grand\_summary\_row value 124 grand\_summary\_row\_border\_width TRUE grand\_summary\_row px 125 grand\_summary\_row\_border\_color TRUE grand\_summary\_row value 126 footnotes\_font\_size TRUE footnotes px 127 footnotes\_padding TRUE footnotes px 128 footnotes\_padding\_horizontal TRUE footnotes px 129 footnotes\_background\_color TRUE footnotes value 130 footnotes\_margin TRUE footnotes px 131 footnotes\_border\_bottom\_style TRUE footnotes value 132 footnotes\_border\_bottom\_width TRUE footnotes px 133 footnotes\_border\_bottom\_color TRUE footnotes value 134 footnotes\_border\_lr\_style TRUE footnotes value 135 footnotes\_border\_lr\_width TRUE footnotes px 136 footnotes\_border\_lr\_color TRUE footnotes value 137 footnotes\_marks FALSE footnotes values 138 footnotes\_spec\_ref FALSE footnotes values 139 footnotes\_spec\_ftr FALSE footnotes values 140 footnotes\_multiline FALSE footnotes logical 141 footnotes\_sep FALSE footnotes value 142 source\_notes\_padding TRUE source\_notes px 143 source\_notes\_padding\_horizontal TRUE source\_notes px 144 source\_notes\_background\_color TRUE source\_notes value 145 source\_notes\_font\_size TRUE source\_notes px 146 source\_notes\_border\_bottom\_style TRUE source\_notes value 147 source\_notes\_border\_bottom\_width TRUE source\_notes px 148 source\_notes\_border\_bottom\_color TRUE source\_notes value 149 source\_notes\_border\_lr\_style TRUE source\_notes value 150 source\_notes\_border\_lr\_width TRUE source\_notes px 151 source\_notes\_border\_lr\_color TRUE source\_notes value 152 source\_notes\_multiline FALSE source\_notes logical 153 source\_notes\_sep FALSE source\_notes value 154 row\_striping\_background\_color TRUE row value 155 row\_striping\_include\_stub FALSE row logical 156 row\_striping\_include\_table\_body FALSE row logical 157 container\_width FALSE container px 158 container\_height FALSE container px 159 container\_padding\_x FALSE container px 160 container\_padding\_y FALSE container px 161 container\_overflow\_x FALSE container overflow 162 container\_overflow\_y FALSE container overflow 163 ihtml\_active FALSE interactive logical 164 ihtml\_height FALSE interactive px 165 ihtml\_use\_pagination FALSE interactive logical 166 ihtml\_use\_pagination\_info FALSE interactive logical 167 ihtml\_use\_sorting FALSE interactive logical 168 ihtml\_use\_search FALSE interactive logical 169 ihtml\_use\_filters FALSE interactive logical 170 ihtml\_use\_resizers FALSE interactive logical 171 ihtml\_use\_highlight FALSE interactive logical 172 ihtml\_use\_compact\_mode FALSE interactive logical 173 ihtml\_use\_text\_wrapping FALSE interactive logical 174 ihtml\_use\_page\_size\_select FALSE interactive logical 175 ihtml\_page\_size\_default FALSE interactive values 176 ihtml\_page\_size\_values FALSE interactive values 177 ihtml\_pagination\_type FALSE interactive value 178 page\_orientation FALSE page value 179 page\_numbering FALSE page logical 180 page\_header\_use\_tbl\_headings FALSE page logical 181 page\_footer\_use\_tbl\_notes FALSE page logical 182 page\_width FALSE page value 183 page\_height FALSE page value 184 page\_margin\_left FALSE page value 185 page\_margin\_right FALSE page value 186 page\_margin\_top FALSE page value 187 page\_margin\_bottom FALSE page value 188 page\_header\_height FALSE page value 189 page\_footer\_height FALSE page value 190 quarto\_disable\_processing FALSE quarto logical 191 quarto\_use\_bootstrap FALSE quarto logical 192 latex\_use\_longtable FALSE latex logical 193 latex\_tbl\_pos FALSE latex value  value 1 NA 2 NA 3 auto 4 fixed 5 auto 6 auto 7 #FFFFFF 8  9 system-ui, Segoe UI, Roboto, Helvetica, Arial, sans-serif, Apple Color Emoji, Segoe UI Emoji, Segoe UI Symbol, Noto Color Emoji 10 16px 11 normal 12 normal 13 #333333 14 #FFFFFF 15 TRUE 16 solid 17 2px 18 #A8A8A8 19 none 20 2px 21 #D3D3D3 22 TRUE 23 solid 24 2px 25 #A8A8A8 26 none 27 2px 28 #D3D3D3 29 NA 30 center 31 125% 32 initial 33 85% 34 initial 35 4px 36 5px 37 solid 38 2px 39 #D3D3D3 40 none 41 1px 42 #D3D3D3 43 NA 44 100% 45 normal 46 inherit 47 5px 48 5px 49 none 50 1px 51 #D3D3D3 52 solid 53 2px 54 #D3D3D3 55 solid 56 2px 57 #D3D3D3 58 none 59 1px 60 #D3D3D3 61 FALSE 62 {1}, {2} 63 NA 64 100% 65 initial 66 inherit 67 8px 68 5px 69 solid 70 2px 71 #D3D3D3 72 none 73 1px 74 #D3D3D3 75 solid 76 2px 77 #D3D3D3 78 none 79 1px 80 #D3D3D3 81 NA 82 FALSE 83 solid 84 1px 85 #D3D3D3 86 none 87 1px 88 #D3D3D3 89 solid 90 2px 91 #D3D3D3 92 solid 93 2px 94 #D3D3D3 95 8px 96 5px 97 NA 98 100% 99 initial 100 inherit 101 solid 102 2px 103 #D3D3D3 104 5px 105 NA 106 100% 107 initial 108 inherit 109 solid 110 2px 111 #D3D3D3 112 8px 113 5px 114 NA 115 inherit 116 solid 117 2px 118 #D3D3D3 119 8px 120 5px 121 NA 122 inherit 123 double 124 6px 125 #D3D3D3 126 90% 127 4px 128 5px 129 NA 130 0px 131 none 132 2px 133 #D3D3D3 134 none 135 2px 136 #D3D3D3 137 numbers 138 ^i 139 ^i 140 TRUE 141  142 4px 143 5px 144 NA 145 90% 146 none 147 2px 148 #D3D3D3 149 none 150 2px 151 #D3D3D3 152 TRUE 153  154 rgba(128,128,128,0.05) 155 FALSE 156 FALSE 157 auto 158 auto 159 0px 160 10px 161 auto 162 auto 163 FALSE 164 auto 165 TRUE 166 TRUE 167 TRUE 168 FALSE 169 FALSE 170 FALSE 171 FALSE 172 FALSE 173 TRUE 174 FALSE 175 10 176 10, 25, 50, 100 177 numbers 178 portrait 179 FALSE 180 FALSE 181 FALSE 182 8.5in 183 11.0in 184 1.0in 185 1.0in 186 1.0in 187 1.0in 188 0.5in 189 0.5in 190 FALSE 191 FALSE 192 FALSE 193 !t  $`\_transforms` list()  $`\_locale` $`\_locale`$locale NULL   $`\_has\_built` [1] FALSE  attr(,"class") [1] "gt\_tbl" "list" |

Table 2 presents the Pearson correlation coefficients between CD4 immune markers and body composition variables. CD4+ count showed a modest positive correlation with both CD4 immune activation count (r = 0.22) and fat mass (r = 0.17), and a negative correlation with lean body mass (r = –0.18). Notably, fat in kilograms was strongly negatively correlated with lean body mass (r = –0.94), indicating an inverse relationship between fat and lean tissue. Participant age showed very weak correlations with all variables, suggesting age may not substantially confound relationships between immune activation and body composition in this sample.

## 4.2 Basic statistical analysis

We employed a simple linear model to investigate the potential association between our primary outcome, CD4 immune activation count, and our main predictor variable, sex. To enhance the model’s explanatory power, we incorporated additional predictors that might influence immune activation counts through multivariate linear regression analysis.

Given that fat content serves as a contributing factor to immune activation in disease progression, individuals with higher fat percentages typically demonstrate reduced susceptibility to tuberculosis, and vice versa. We analyzed the distribution of fat composition between men and women. This (**plot2-result?**) presents a boxplot illustrating the distribution of fat in kilograms across male and female participants.



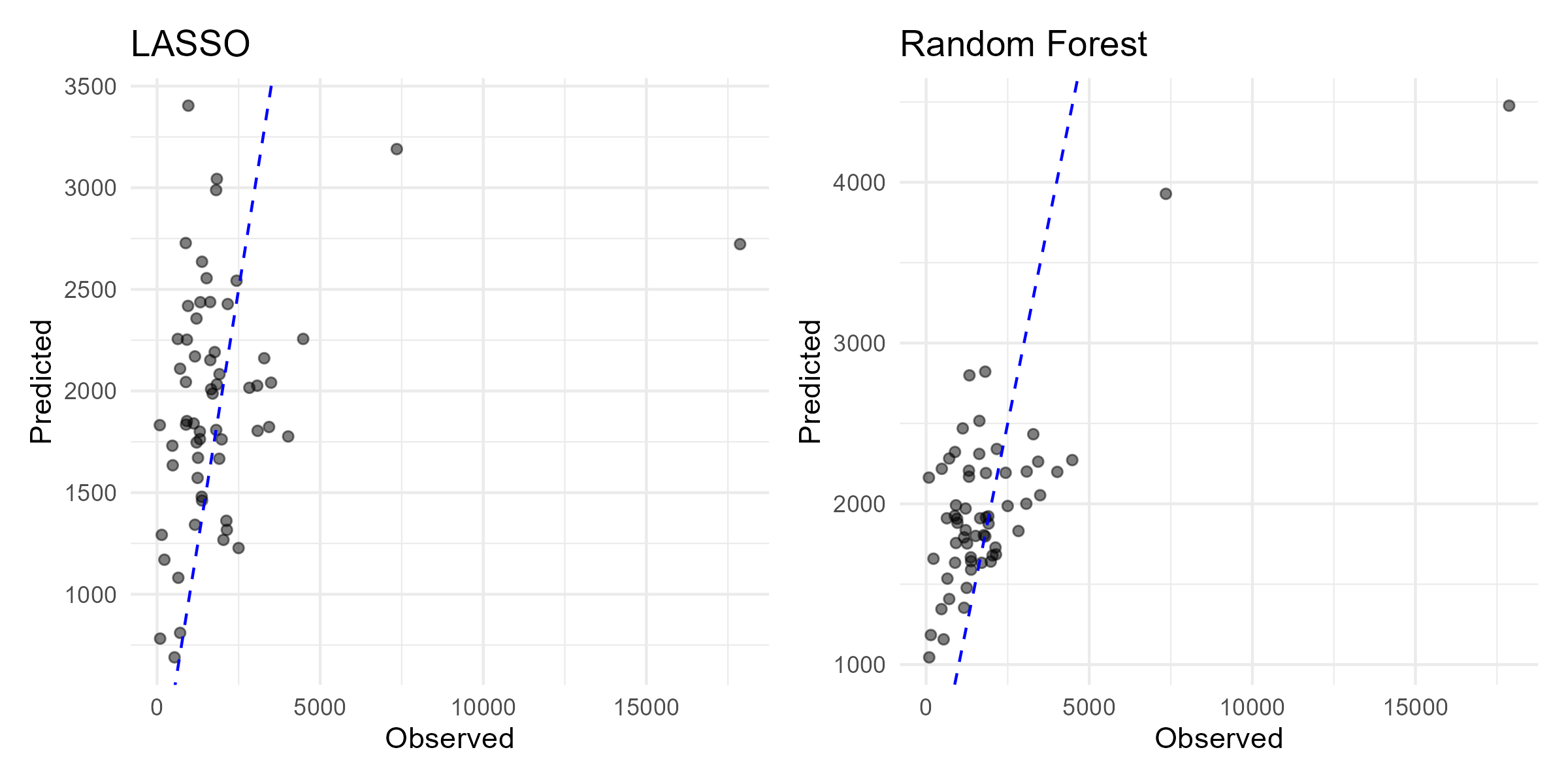
Box plot showing Fat in kg among males and females.

## 4.3 Full analysis

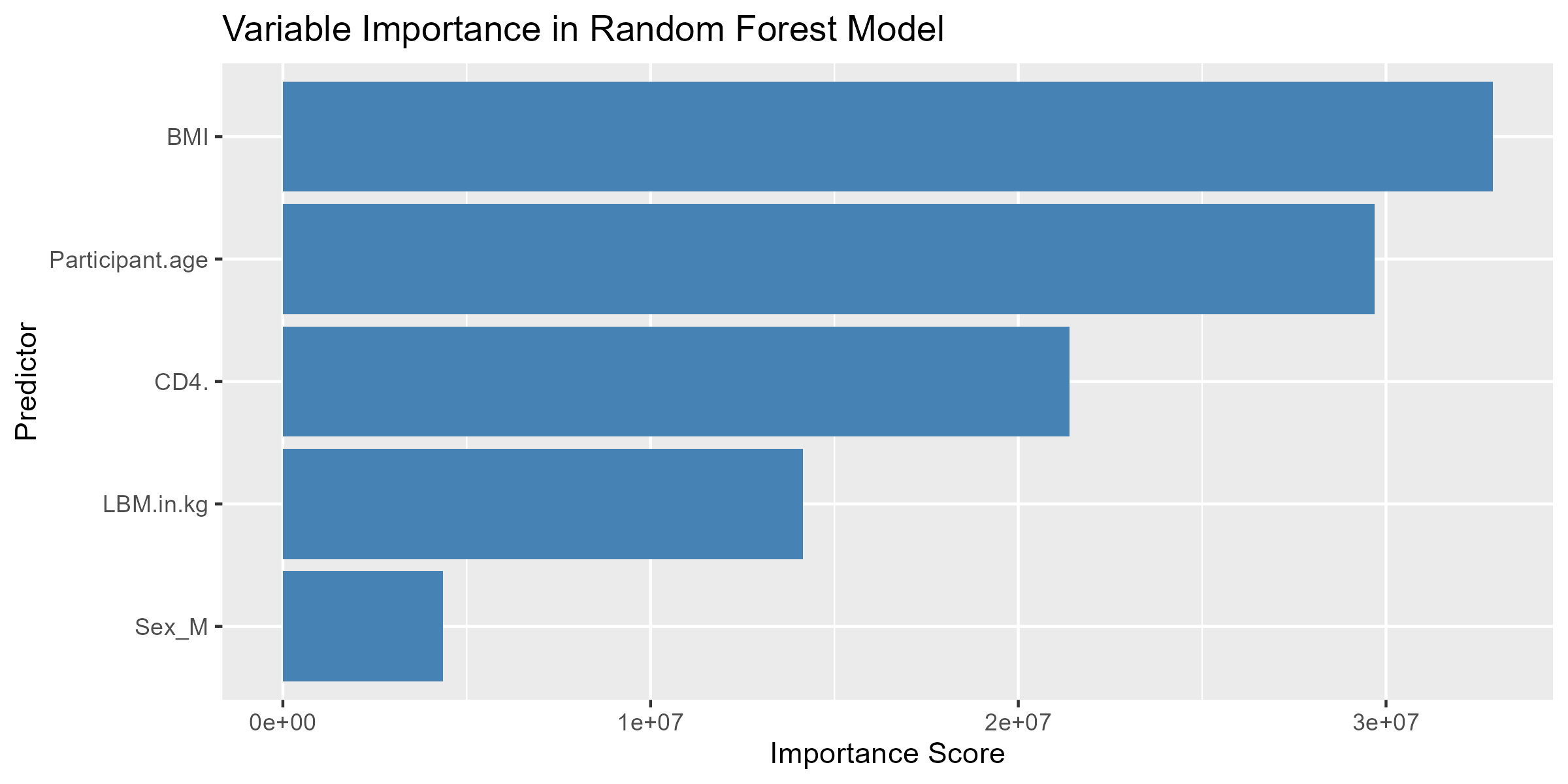
In order to take care of confounding, we performed a multi-regression analysis and the results of the analysis were not satifactory.

We finally considered using ML algorithm to fit our data. The Random Forest model achieved the lowest RMSE (1132), outperforming both the linear regression and LASSO models, which had RMSEs around 2107. This figure (**Predicted\_vs\_observed?**) shows a comparison of observed vs. predicted values for each model.

This figure (**Variable\_plot?**) shows the variable importance plot from the random forest model. This highlights the most influential predictors of CD4 immune activation. BMI, participants age, CD4+, Lean body mass(LBM) and Sex emerged as the most important features, indicating they contributed most to the model’s predictive accuracy. This suggests that body composition and biological sex may play a key role in immune activation patterns.



predicted vs observed of three models



Variable plot from the random forests model.

We applied LASSO regression to identify predictors of CD4 immune activation. After removing the participant ID variable, the model retained BMI (estimate = -297), CD4+ count (estimate = 514), sex (male) (estimate = 260), and lean body mass (estimate = -22.7) as non-zero coefficients. Participant age was excluded from the final model.

Interestingly, LBM (Lean Body Mass) and Participant Age were ranked lower, suggesting they had less influence in this dataset. These findings align with the Random forest results, which also retained BMI,CD4+ lean body mass and sex as key predictors.

# 5. Discussion

In this analysis of sex-based differences in TB-related immune activation, we aimed to identify key predictors of CD4 immune activation count , which we used as a proxy for severity of disease. Our hypothesis was that higher immune activation counts indicate more severe illness, and that biological sex may influence immune responses to TB.

when we used both the LASSO regression and random forest models, we identified CD4+ count, BMI, sex, and lean body mass (LBM) as important predictors of immune activation. Our findings highlight a notable difference in CD4 immune activation count rates between males and females, with males exhibiting higher risk of TB infection and disease than females. Sex was retained as a significant predictor, with a positive coefficient (estimate = 260). This indicated that males experienced a 260 increase in CD4 immune activation count compared to females, who served as the reference group. This suggests that males may experience more severe immune activation, supporting the hypothesis that sex-based biological differences contribute to disease severity in TB. This result aligns with previous studies that have reported sex-based disparities in health outcomes and mortality risks as a result of TB(7),(8). Some studies still attribute the increased risk among males to be as a result of biological differences or differences in the accessibility of resources, especially healthcare. This is generally because women have better prognosis than men and they have better health seeking behavior as compared to their male counterparts.

Research shows that most women have higher fat repositories than men(9) and some studies show that the higher the fat percentage an individual has, the lower the risk of progression of the tuberculosis infection to disease hence lowering mortality rates among females compared to men. In addition, women tend to have more lean body mass compared to men hence having a lower chance of suffering from adverse effects of TB compared to men.

## 5.1 Strengths and Limitations

This study has several strengths. We applied simple linear regression, LASSO regression and random forest models for annalysis of our data. We were able to capture a range of relationships between predictors and immune activation, which in turn increased the robustness of our findings. The focus on sex-based differences in TB immune response addressed an important and often neglected question in infectious disease research especially among TB patients who are HIV negative. Additionally, the use of biologically meaningful predictors such as BMI, lean body mass, CD4+ count, and sex enhances the interpretability and relevance of the results. The consistency of key predictors across the LASSO model and the random forests model further supports the reliability of our conclusions. However, there are limitations to consider. The relatively small sample size may reduce the generalizability and stability of the findings. As an observational study, we are limited to identifying associations rather than making causal claims. Important clinical variables, such as TB stage or treatment history, may not have been included, potentially introducing residual confounding. Moreover, while CD4 immune activation was used as our marker for disease severity, it may not fully capture the clinical picture.

Despite limitations of the present of our study, some studies show that Patients with wasting do not recover to the same levels of body mass as those who were not wasted. This suggests that tuberculosis may leads to permanent loss of lean tissue and fat mass.

## 5.2 Conclusions

Our results support the hypothesis that higher immune activation is associated with greater disease severity, and demonstrate that males exhibit higher CD4 immune activation levels compared to females, indicating they may experience more intense immune responses and possibly more severe disease outcomes. These findings highlight the importance of considering sex-based biological differences in TB research and clinical management.

In addition to sex, BMI and lean body mass emerged as important predictors, suggesting that body composition may influence the immune landscape in TB. These insights highlight the need for sex-specific approaches in TB monitoring and treatment, and point to CD4 immune activation as a potentially valuable biomarker for assessing disease severity. However, the need for future research is needed to validate these findings especially among larger populations to explore the underlying sex mechanism and body composition differences in regard to immune activation.

# 6. References

1. Santos M de LSG, Vendramini SHF, Gazetta CE, et al. Poverty: Socioeconomic characterization at tuberculosis. *Revista Latino-Americana de Enfermagem* [electronic article]. 2007;15(spe):762–767. (<http://dx.doi.org/10.1590/S0104-11692007000700008>)

2. Martinez L, Shen Y, Mupere E, et al. Transmission of mycobacterium tuberculosis in households and the community: A systematic review and meta-analysis. *American Journal of Epidemiology* [electronic article]. 2017;185(12):1327–1339. (<http://dx.doi.org/10.1093/aje/kwx025>)

3. Dye C, Garnett GP, Sleeman K, et al. Prospects for worldwide tuberculosis control under the WHO DOTS strategy. *The Lancet* [electronic article]. 1998;352(9144):1886–1891. (<http://dx.doi.org/10.1016/S0140-6736(98)03199-7>)

4. Murgia Y, Sepulcri C, Crupi L, et al. Tuberculosis infection control: Experiences and considerations from a web based tool implementation. In: *pHealth 2024*. IOS Press; 2024(<http://dx.doi.org/10.3233/SHTI240053>)

5. Whalen C. Sex differences in the presentation of tuberculosis in uganda. 2023;(<https://dataverse.unc.edu/citation?persistentId=doi:10.15139/S3/AYOFEU>)

6. Horton KC, MacPherson P, Houben RMGJ, et al. Sex differences in tuberculosis burden and notifications in low- and middle-income countries: A systematic review and meta-analysis. *PLOS Medicine* [electronic article]. 2016;13(9):e1002119. (<http://dx.doi.org/10.1371/journal.pmed.1002119>)

7. Hertz D, Schneider B. Sex differences in tuberculosis. *Seminars in Immunopathology* [electronic article]. 2018;41(2):225–237. (<http://dx.doi.org/10.1007/s00281-018-0725-6>)

8. Marçôa R, Ribeiro AI, Zão I, et al. Tuberculosis and gender – factors influencing the risk of tuberculosis among men and women by age group. *Pulmonology* [electronic article]. 2018;24(3):199–202. (<http://dx.doi.org/10.1016/j.pulmoe.2018.03.004>)

9. Karastergiou K, Smith SR, Greenberg AS, et al. Sex differences in human adipose tissues – the biology of pear shape. *Biology of Sex Differences* [electronic article]. 2012;3(1). (<http://dx.doi.org/10.1186/2042-6410-3-13>)