

# Results, conclusion and discussion of EDA

Rienk Heins

9/26/2020

## Results

One of the things to looked at is if age or gender has any impact on the difference in allergic reaction. A difference in gender might not be expected but age could have an impact since in most cases health will deteriorate as a person gets older since the immune system will get weaker with age.

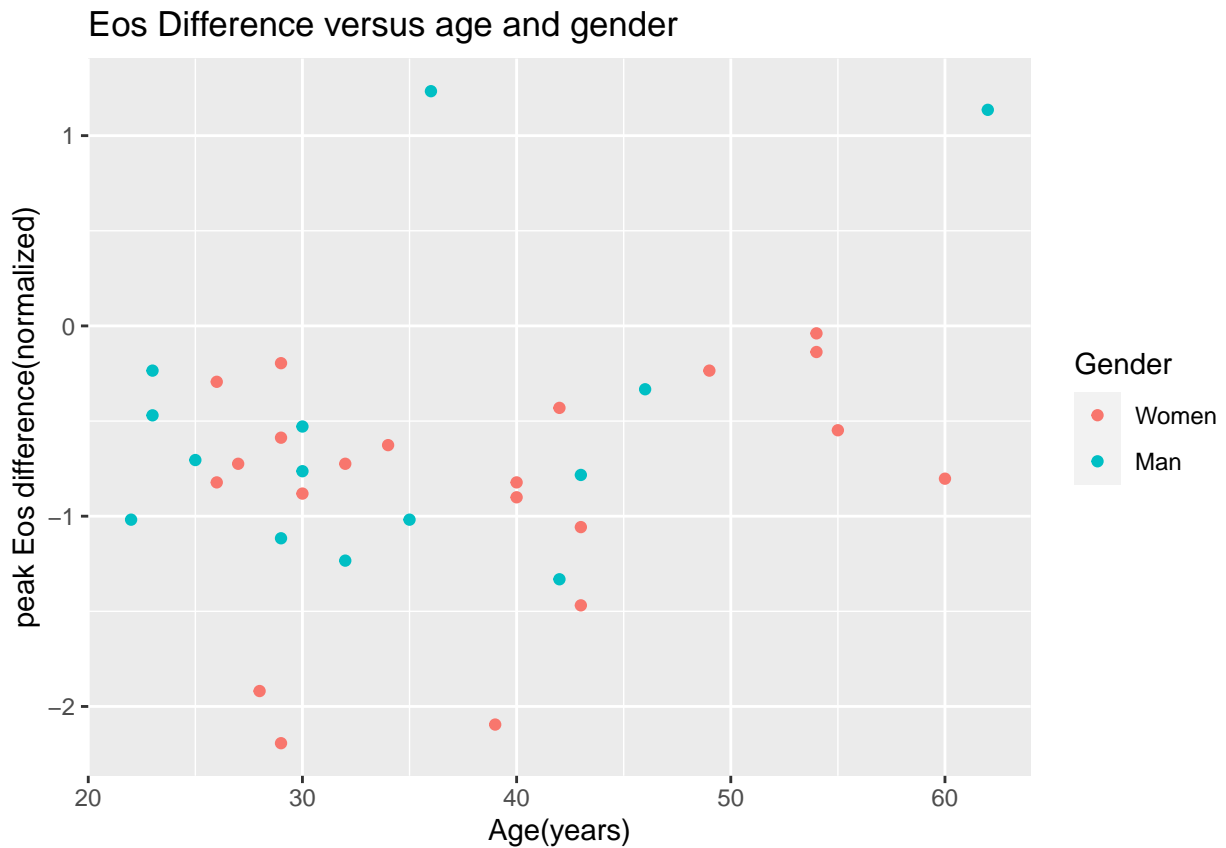


Figure 1: Eos difference at different ages and gender

As seen in figure 1 gender doesn't seem to impact the reaction difference a lot. Even though all negative outliers are female and all positive outliers male, most male and female points fall between the same values. But when looking at age there seems to be an effect as when the age keeps going up the decrease in allergic reaction gets lower, even though in most cases there is still a decrease for older patients the effect of the diet changes seem to get less impactful as age gets higher with only one patient getting a decrease higher than 0,7.

Another variable looked at is the fat intake and the direct impact of this intake on allergic reactions, so not the change in reaction. This is looked into as fats are required for hormones which in regulate the body, so

an imbalance in hormones might effect the allergic response and could be caused by a lack of fat intake. A second variable looked at is the intake of saturated fat as this fat type is generally seen as bad for a persons health.

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

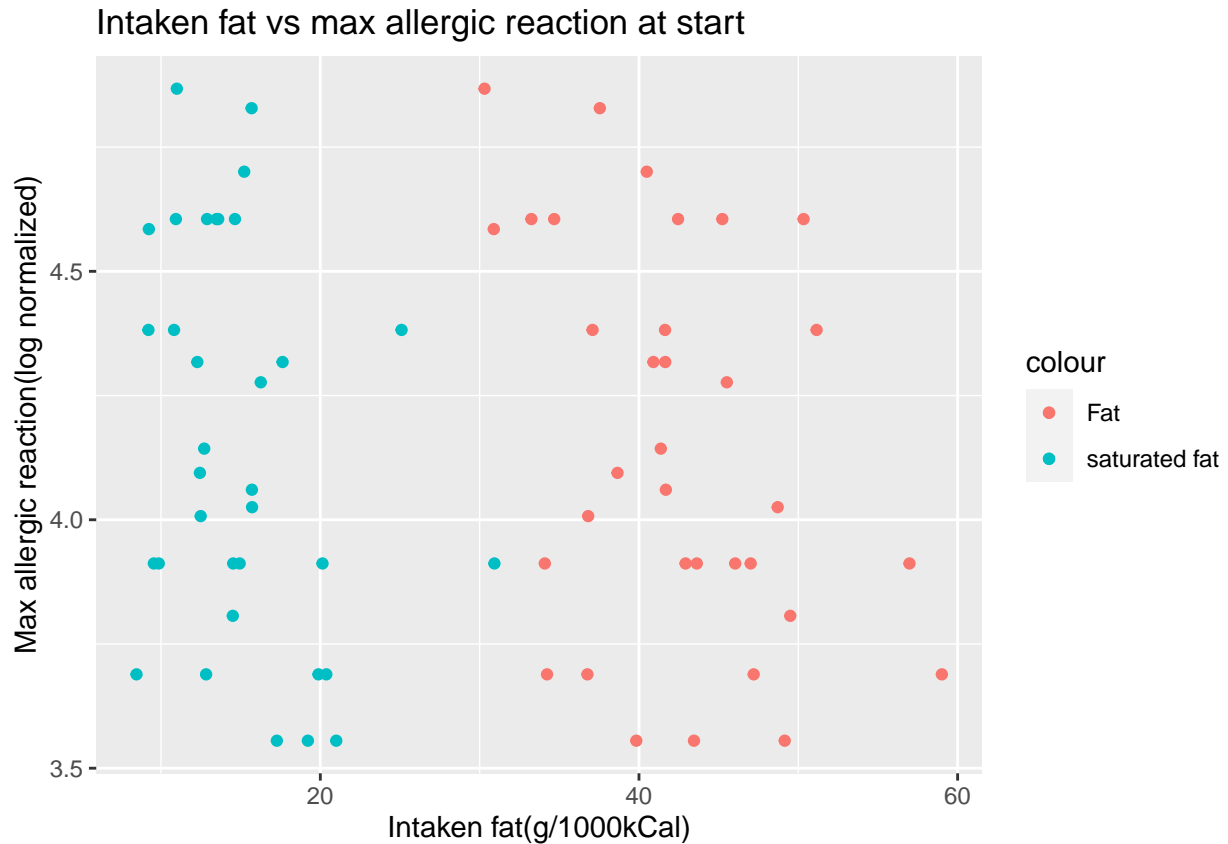


Figure 2: Effect of fat intake on alllergic effect at start

A thing to note at figure 2 at the beginning of the experiment is that as the fat intake increases, the allergic reaction gets lower. This could have been expected as fat intake has direct impact on hormonal balance, most notably that the highest peaks of allergic reaction are at the lowest fat intake levels. A similar thing somewhat happens with saturated fat as there are no peaks of allergic reactions at higher levels of saturated fats, but at the lower values the points are quite evenly distributed over all levels of allergic reaction, so a high intake of saturated fat might lower allergic response, which is against the first stated hypothesis. Yet there aren't a lot of patients with an intake of over 20g/1000kCal so this isn't heavily supported.

A notable thing from figure 3 in contrast in figure 2 is that saturated fat increase doesn't seem to effect the allergic reaction as much anymore, since there are no peaks after about 15g/1000kCal, there isn't a big decline anymore as seen in figure 2. Something that figure 3 does seem to have in common with figure 2 is that at lower fat intake there seem to be mostly higher allergic reactions. Yet after six weeks there isn't an incline at higher fat intakes as seen in figure 2, after an intake of 50 g/1000kCal most reactions are above 3. So this seems to support that there is a minimum amount of fat intake that gives a decrease in allergic reactions, but as the intake gets higher this doesn't seem to positively effect the allergic reaction.

Next variables looked at where the macro nutrients and the difference in intake of there molecules on the change in allergic reaction. As these nutrients are the bulk of food intake in terms of pure weight and might

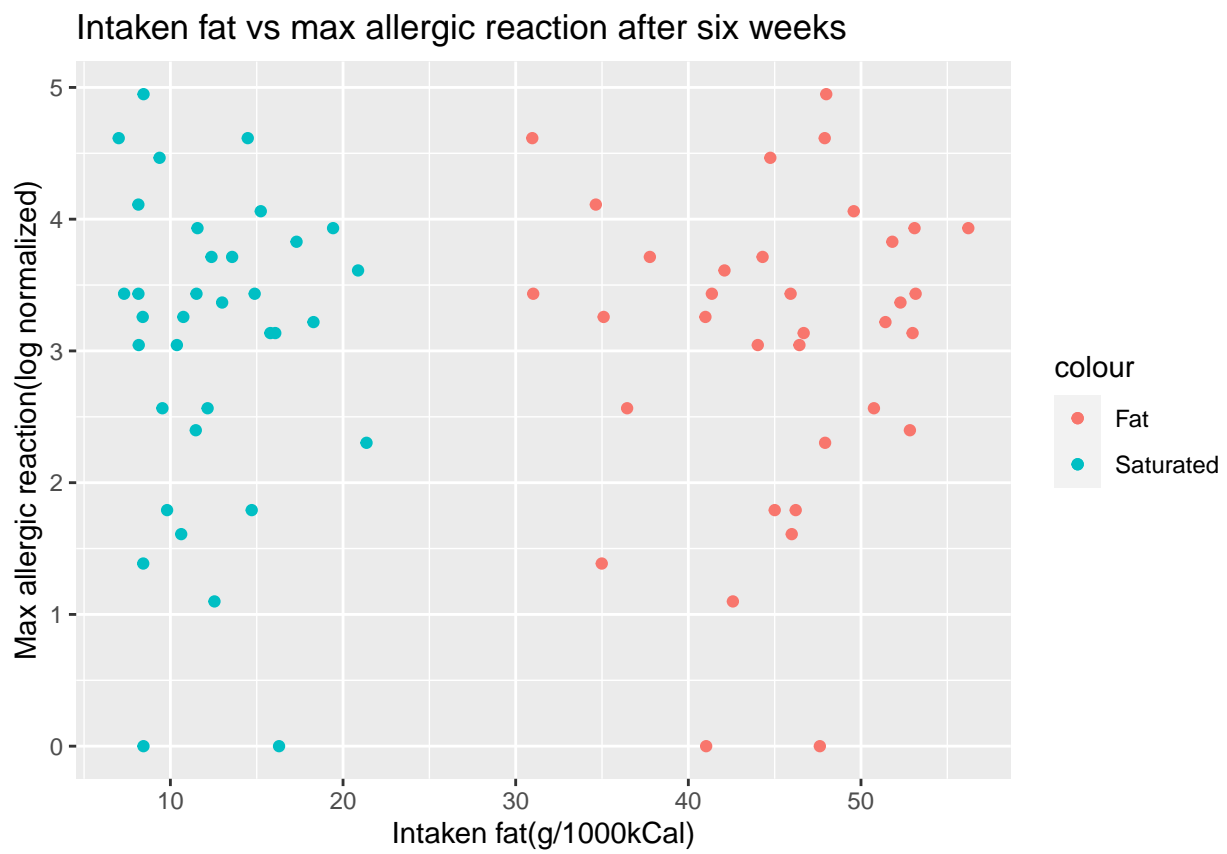


Figure 3: Effect of fat intake on alllergic effect at after 6 weeks

thus impact the allergic reaction.

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

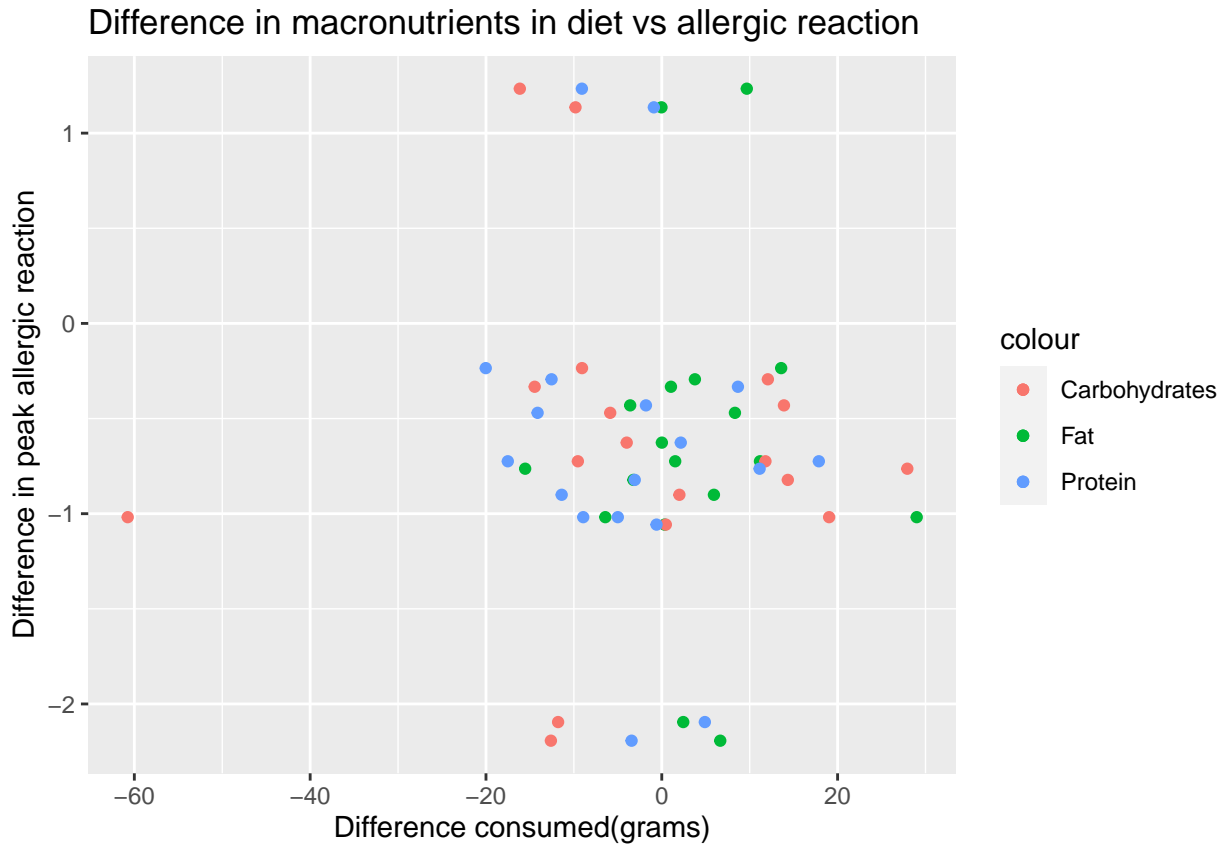


Figure 4: Effect of macro nutrients on allergic reaction

One thing in figure 4 contradicting figures 2 and 3 is that lowering in fat seems to give a bigger decrease in immune response as for an increase the points are relatively evenly distributed in above and under a decrease of 0,5, at a decrease of fat intake there is only one under this value and the rest above. So a decrease in fat seems to decrease the allergic reaction more, as long as the fat intake stays above a certain value. In terms of carbohydrates, the biggest decreases in allergic reaction happen at an increase in carbohydrates. This also might be the case for protein as with a decrease there is a more even distribution but at an increase only one of the four decreases is under 0,5, yet there are only four instances of protein increase, not counting the far outliers as they seem to be incorrect, thus there isn't a lot of support for this claim.

One of the variables looked at because of high correlation is retinol activity equivalents(RAE), which accounts for the different bioactivities of retinol and provitamin A carotenoids, which all eventually will be converted to retinol.

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

As seen in figure 5 RAE doesn't seem to have a lot of impact on the allergic reactions as most points are around the same values in peak Eos at different changes in RAE with the exception of some extreme outliers.

The next variable looked at is difference in calcium intake as it also gave a high score on a correlation test.

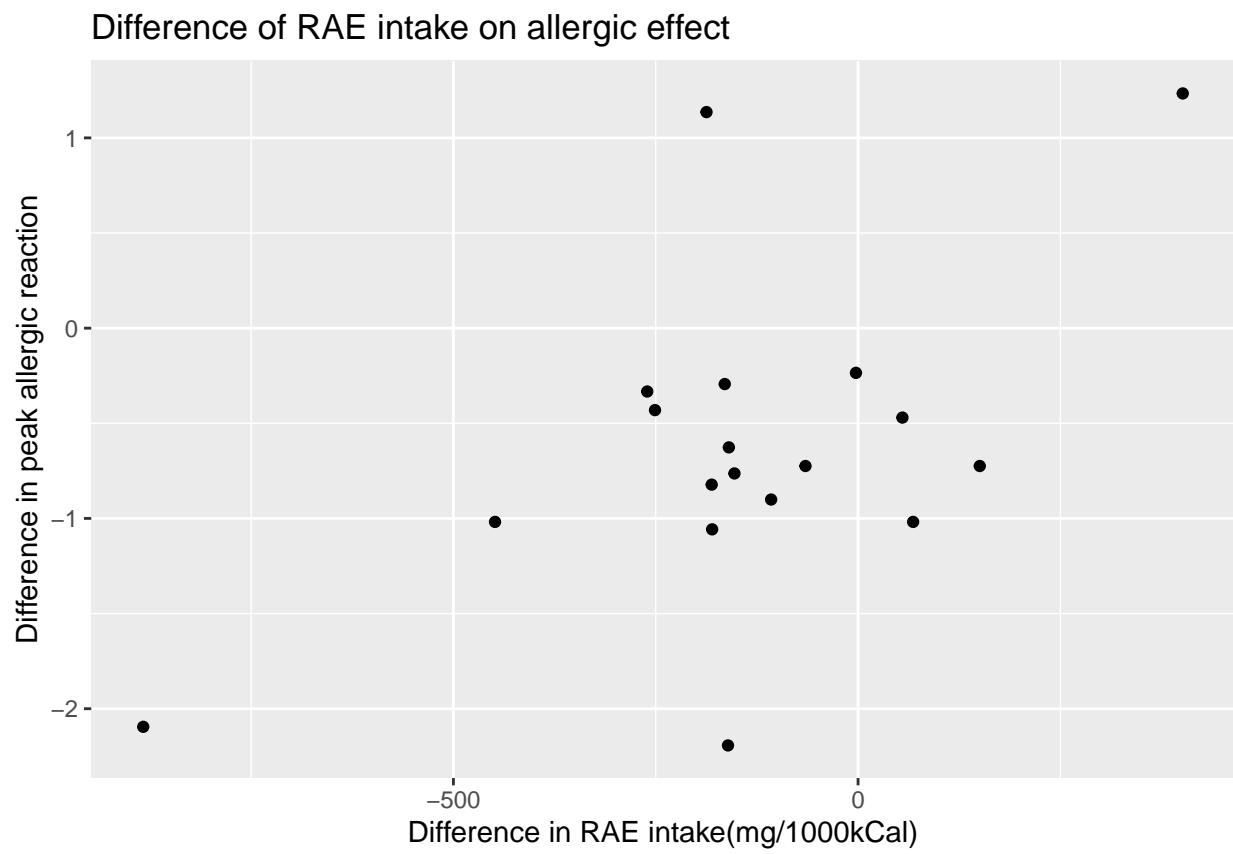


Figure 5: Effect of retinol activity equivalents(RAE)

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

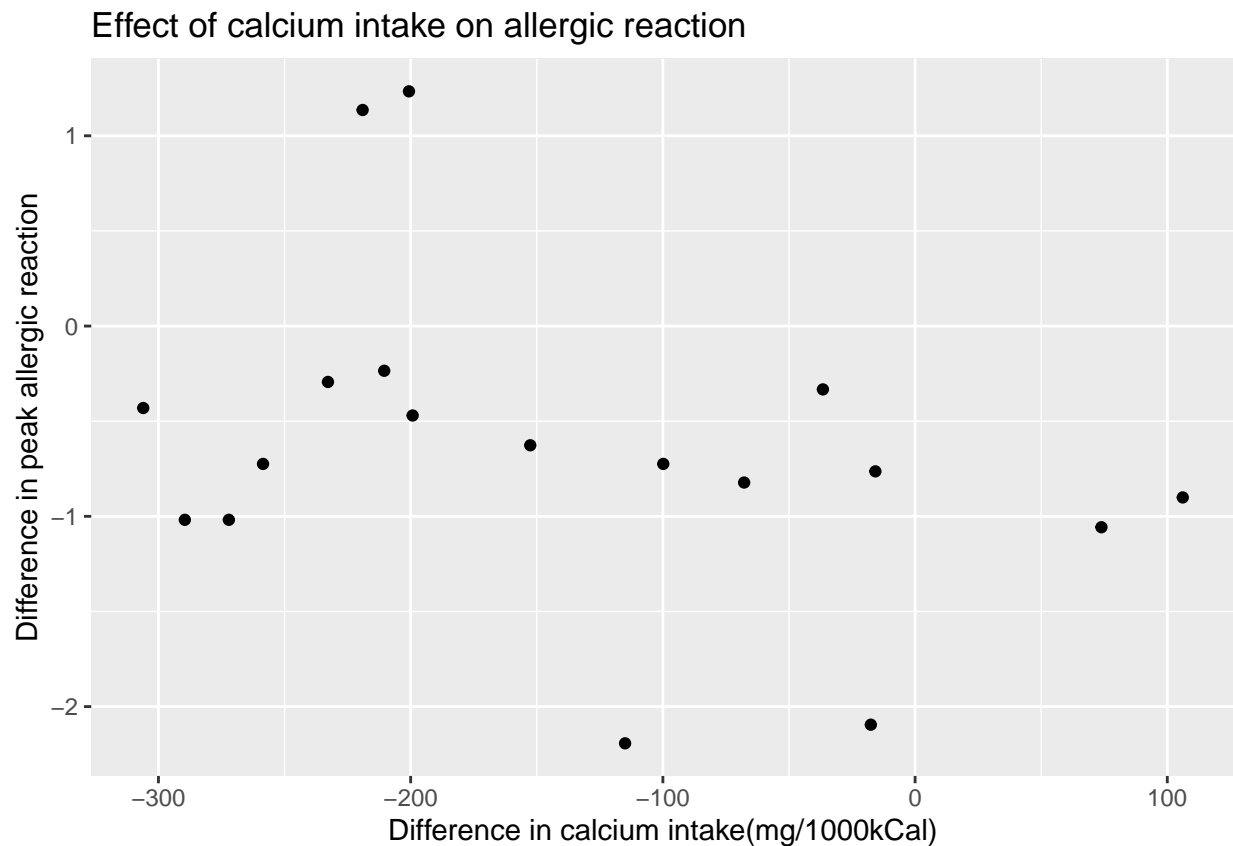


Figure 6: Effect of calcium difference on allergic reactions

The overall trend of the graph in figure 6 is that as calcium intake increases the allergic effect reduces with higher decreases in calcium giving lower decreases in reaction with the exception at a high decrease of more than 250 mg/1000kCal.

The next variable that was visualized was the difference in vitamin C intake as it gave a high score in the correlation test and as it boosts the immune system, an effect which might help in lowering the allergic reaction.

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

As seen in figure 7 the decrease in allergic reaction increases as the vitamin C intake gets higher, with a exception at a low decrease to around 14 mg/1000kCal max.

The last variable that was looked at is the mineral zinc, as it also scored high on the correlation test with difference in allergic reaction.

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

As seen in figure 8 the effect of zinc doesn't seem to impact the allergic reactions to much as the same points are found at different zinc differences and there doesn't seem to be effect on the increase or decrease of zinc.

## Conclusion

The question asked in the beginning was:

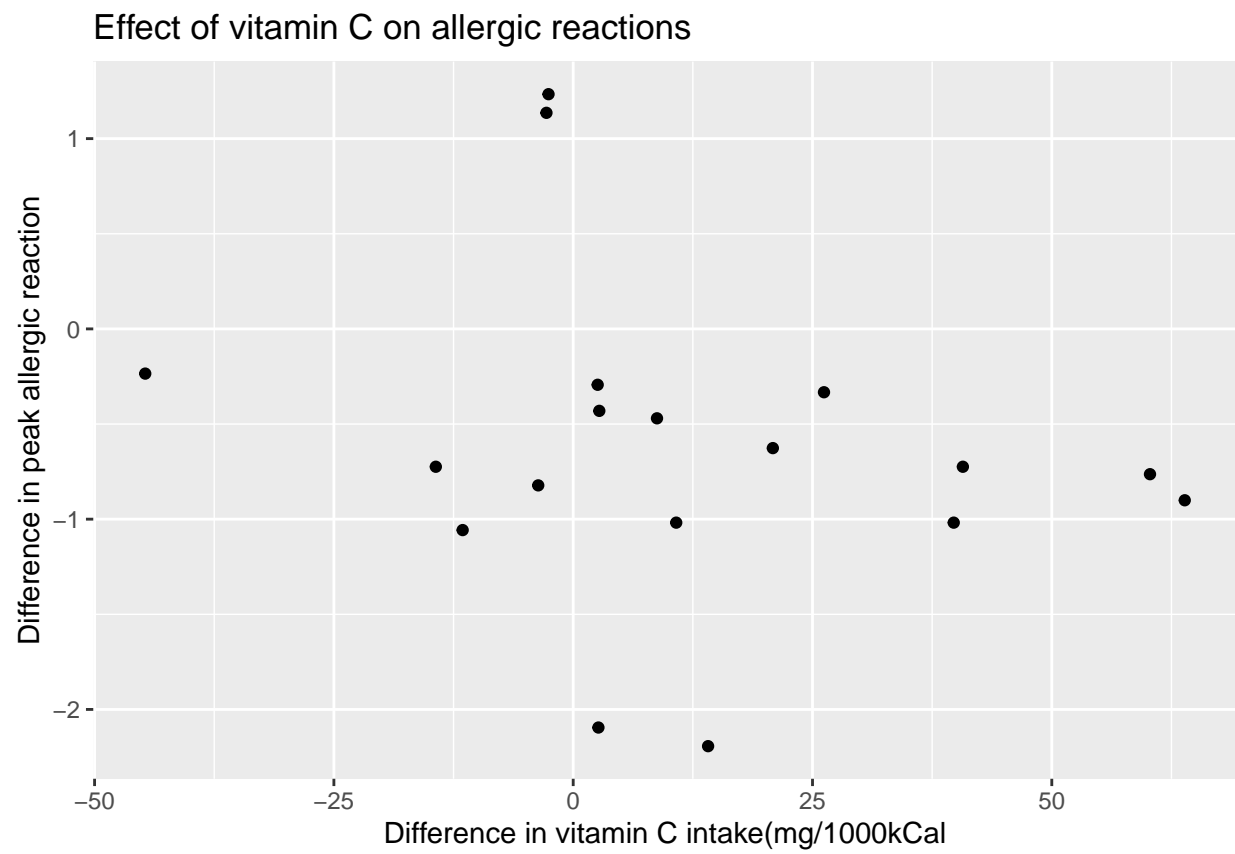


Figure 7: Effect of vitaminC difference on allergic reaction

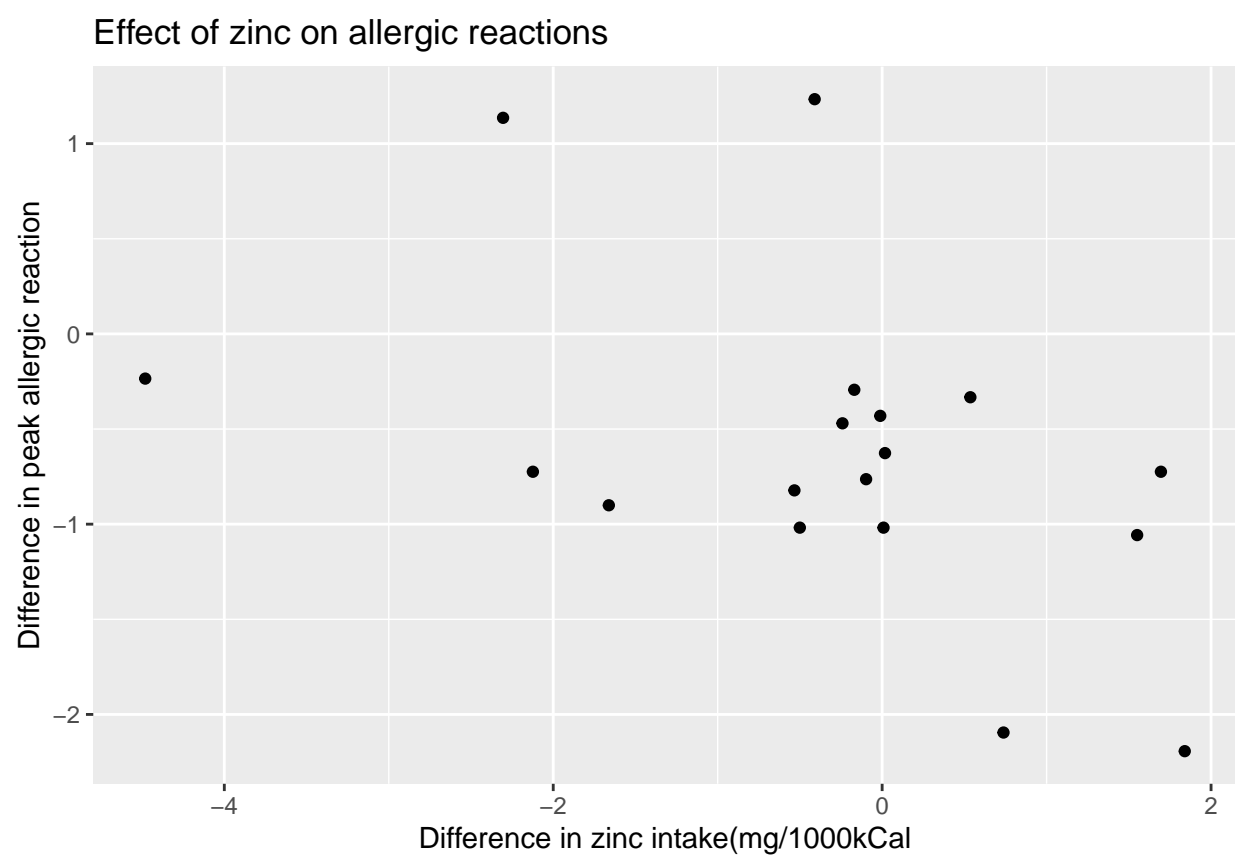


Figure 8: Effect of zinc difference on allergic reactions



Can the allergic effect of eosinophilic esophagitis in a patient be predicted using machine learning with the dietary composition of foods and nutrients of this patient?

With the data it seems that in order to lower the allergic reaction in a patient a diet should be made containing a minimum amount of fat but with a decrease in fat consumed with an increase in protein, carbohydrates, vitamin C and calcium. Yet there doesn't seem to be a variable that gives significantly higher effects than others so in order to create a machine learning program to predict the difference in allergic reaction all these variables and more will need to be accounted for.

## **Discussion**

There are still some points to discuss. First of all like said before there is a lot of data missing because the original study isn't finished as writing of this text. Because of this a lot of variables that will impact the end results cannot be examined and used. So the end results of this research might not be correct because not all of the variables impacting the allergic reaction are known. Also the test group is quite small because of this, a total of only 40 patients with data of the precise difference in nutrients of only 18 patients. A bigger group would give more certainty as if effect seen in the figures are correct or it they might not have an as big effect as thought.

So for a reproduction of this research it would be recommended to use the full data set of the study if it is presented at that time. Also it would help to bring in more patients for the experiment as more test subjects would give more certainty to information from the figures.

Also a better tracking could have been done on patients that stopped in the middle of the study or have wrong data as there seem to be a few pretty extreme outliers.