BIOL606
Midterm 3
Gavrila Ang

The inflammatory profiles of hypertensive adults in different total cholesterol level groups

METHODS

I categorised a clinical population of 157 hypertensive older adults into 3 groups (Acceptable, Borderline, High) according to their "total cholesterol" serum measurements. These groupings were based on the reference intervals (see Table 1) of the LabCorp test used. I created an extensive inflammatory profile consisting of the following descriptive variables: absolute counts of peripheral neutrophils, monocytes, eosinophils, basophils, lymphocytes, and immature granulocytes, as well as serum levels (pg/mL) of CRP, SAA, soluble ICAM and soluble VCAM (Revelle W, 2022). I generated a correlation matrix for these descriptive variables using pairs.panel() from the psych package and found strong collinearity between the ICAM and VCAM variables (R = 0.77. CRP also approached collinearity with SAA (R = 0.58) and ICAM (R = 0.53). Due to the collinear nature of the dataset, I decided to move forward with a principal component analysis (PCA) to analyse the inflammatory profiles of each clinical group.

Table 1: LabCorp Reference Intervals for Total Cholesterol Serum Measurements

Age (y)	Acceptable	Borderline	High
> 19	<200 (or 100-199)	200-239	240

As the descriptive variables were measured in different units, the data was centred and scaled to allow for normalised comparisons. I performed a PCA, using prcomp() from the base R package (R Core Team, 2021), on the data and generated 10 new principal components (PC) that were non-collinear. The first 6 principal components cumulatively described ~82.3% of the variation in dataset. Using fviz_contrib() from the factoextra package (Kassambara and Mundt, 2020), I visualised the contributions of each variable towards the first 6 principal components (see Figure 1). The dotted red reference line in this bar plot determined which variables provided the greatest contributions to explaining variation within the dataset and would be retained for subsequent re-analysis.

Table 2: Principal Components Derived and Proportion of Variation

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Cumulative Proportion	0.256	0.427	0.541	0.640	0.736	0.823	0.893	0.943	0.984	1.000

Thereon, I performed another PCA on the reduced dataset and generated 4 new principal components. The first 2 principal components successfully described ~71% of the variation in the reduced dataset (see Table 3). Using fviz_pca_biplot() from the factoextra package (Kassambara and Mundt, 2020), I generated biplots with the reduced PCA and labelled the points according to "total cholesterol" category (see Figure 2). Note that the biplots are labelled Dim1 and Dim2, corresponding to PC1 and PC2, respectively.

Figure 1: Contribution scores of descriptive variables to PC1-6

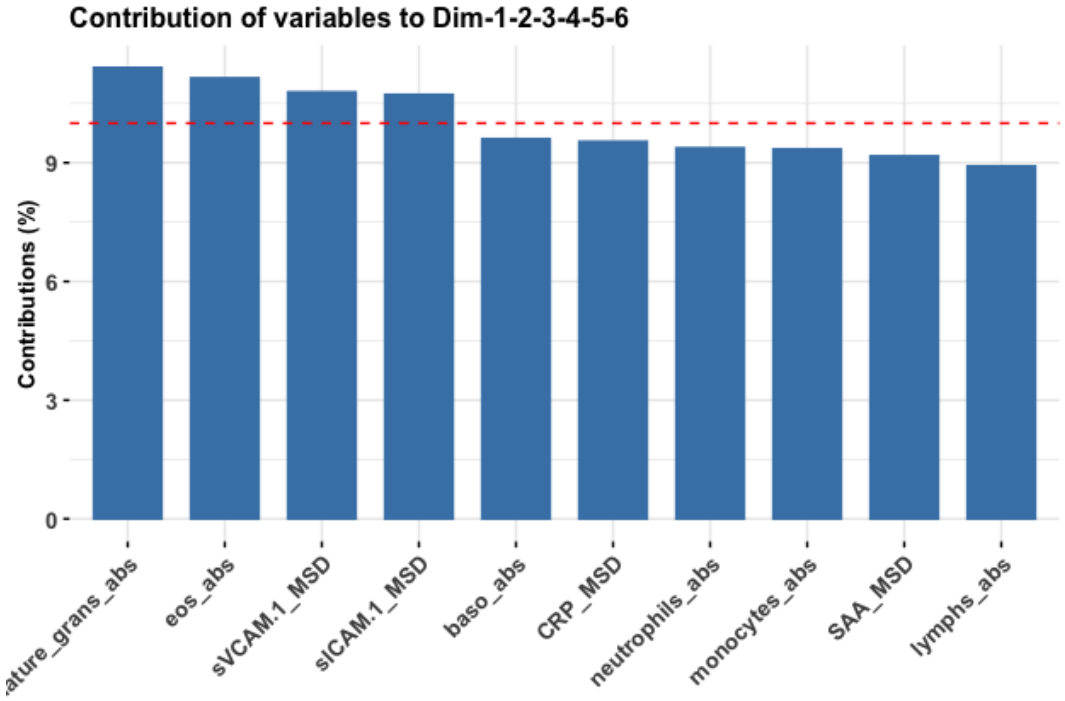


Table 3: Principal Components Derived and Proportion of Variation

	PC1	PC2	PC3	PC4
Cumulative Proportion	0.445	0.714	0.944	1.000

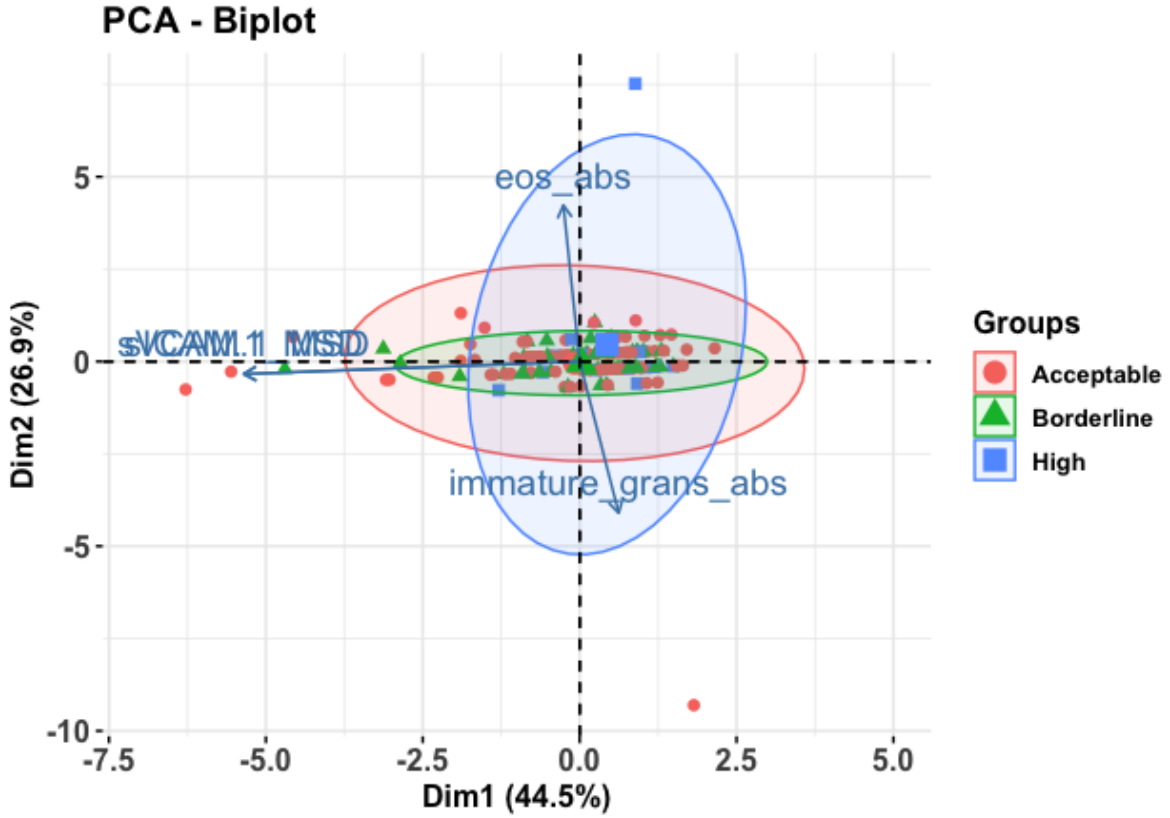
RESULTS

The following descriptive variables: sICAM, sVCAM, eosinophils, and immature granulocytes were found to describe most of the variation within the dataset (see Figure 2). The highly collinear nature of soluble sICAM and sVCAM was reflected on the biplot, as their respective arrows overlapped (no angle between). As such, it can be derived that the two variables can be used as proxies for each other in future investigations.

Despite the high variation described by first two dimensions in the final PCA analyis, there was no visual evidence of clustering amongst the points. The ellipses drawn (at a 95% confidence interval) around points from the same clinical group, overlap with each other. Moreover, there appeared no progressive increase in affinity for the descriptive variables in accordance to clinical group ranking. For example, the “High” group of points did not show more association with sICAM and sVCAM in comparison to the “Borderline” group of points. A confounding observation is that the “Acceptable” ellipses is larger than the “Borderline” ellipses. This indicates that participants within the “Acceptable” group have higher measures of these inflammatory variables compared to “Borderline” participants. This biological contradiction can be explained by the unbalanced dataset, wherein the number of participants in the “Acceptable” group is almost double that of the “Borderline” group. The “Acceptable”

group would naturally more variation, and a few outlier participants with high sVCAM and sICAM measures appear to enlarge/skew the ellipse.

Figure 2: Biplot of Reduced Principal Component Analysis



Overall, these overlapping ellipses indicate that the inflammatory profile created was not successful in grouping the clinical population in accordance with LabCorp reference levels. These results can also be interpreted as the lack of distinct inflammatory profiles in older adults grouped according to "total cholesterol" serum measurements.

It should be noted that when the contributions of all descriptive variables were assessed (see Figure 1), 6 variables had contributions scores very close to the red reference line. This indicates that these variables did provide a marked contribution in explaining variation in the dataset. The inclusion of more descriptive variables would have explained more variation in the dataset and potentially better differentiated between the "total cholesterol" groups. A larger and more well-balanced sample size would also significantly benefit these analyses, to "control" for outliers. Most crucial however, would be the alternative use of an expertly curated inflammatory profile consisting purely of cytokine measurements. This panel may serve as a far superior descriptor of peripheral inflammation and improve clustering of the participants with their appropriate clinical group.

Table 4: Abbreviations List

Abbreviation	Definition
eos_abs	Absolute counts of eosinophils
immature_grans_abs	Absolute counts of immature granulocytes
sICAM.1_MSD	Soluble Intercellular Adhesion Molecule 1 measured using a Meso Scale Discovery kit

sVCAM.1_MSD	Soluble Vascular cell adhesion protein 1 measured using a Meso Scale Discovery kit
baso_abs	Absolute counts of basophils
CRP_MSD	C-reactive protein measured using a Meso Scale Discovery kit
neutrophils_abs	Absolute counts of neutrophils
monocytes_abs	Absolute counts of monocytes
SAA_MSD	Serum amyloid A measured using Meso Scale Discovery kit
lymphs_abs	Absolute counts of lymphocytes

BIBLIOGRAPHY

1. https://www.labcorp.com/tests/001065/cholesterol-total

2. Revelle W (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.2.3, https://CRAN.R-project.org/package=psych.

3. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

4. Alboukadel Kassambara and Fabian Mundt (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra

5. Nguyen MT, Fernando S, Schwarz N, Tan JT, Bursill CA, Psaltis PJ. Inflammation as a Therapeutic Target in Atherosclerosis. *Journal of Clinical Medicine*. 2019; 8(8):1109. https://doi.org/10.3390/jcm8081109

6. Helmy A, Antoniades CA, Guilfoyle MR, Carpenter KL, Hutchinson PJ. Principal component analysis of the cytokine and chemokine response to human traumatic brain injury. *PLoS One*. 2012;7(6):e39677. doi:10.1371/journal.pone.0039677