Relationship between Kidney Dysfunction and Hypertension

Computational Statistics and Probability

Farhat Jahan (800774640)

Introduction

Background

Hypertension which is commonly known as high blood pressure is a cause of chronic kidney disease (CKD) and cardiovascular risk. Kidney disease is a closely connected health problem which significantly impacts global health. Damaged kidneys struggle to release excess fluid and sodium from the body leading to increased blood volume and elevated blood pressure. Impaired kidneys release more hormones like renin which causes blood vessels to tighten ultimately causing high blood pressure [2][3]. A number of studies has shown that there is an association between increased Creatinine and Albumin levels which is an indication of reduced kidney function and increased blood pressure and cardiovascular risk [4][5]. But very few studies investigated the inverse relationship that impaired kidney can cause high blood pressure levels.

Reduced kidney function is often indicated by elevated serum creatinine levels which leads to fluid retention further causing hypertension. However, the bidirectionality of this approach is still uncertain. I am going to address this gap using statistical analysis on NHANES dataset [1]. Identifying reduced kidney function as a primary causal factor could enable clinicians to develop early detection markers and targeted interventions that address fluid retention and vascular resistance before hypertension manifests. Fluid retention is measured by extracellular fluid. This approach could substantially reduce the burden of hypertension-related complications, such as heart failure and stroke, particularly in populations with preexisting renal impairment.

Methods and Study Aims

Primary study aim

To quantify the effect of impaired kidney function on blood pressure levels. Identify and model the mediating role of fluid retention in the pathway from reduced kidney function to elevated blood pressure. Adjust for potential confounding variables such as age, sex, comorbidities, and lifestyle factors to ensure robust estimation.

Secondary study aim

To develop predictive models that estimate hypertension risk based on creatinine levels (Albumin to Creatinine ratio is taken into account). To investigate whether the relationship varies across demographic subgroups. To assess how early identification of impaired renal function can serve as a preventive strategy for hypertension management.

Hypothesis

Hypothesis:

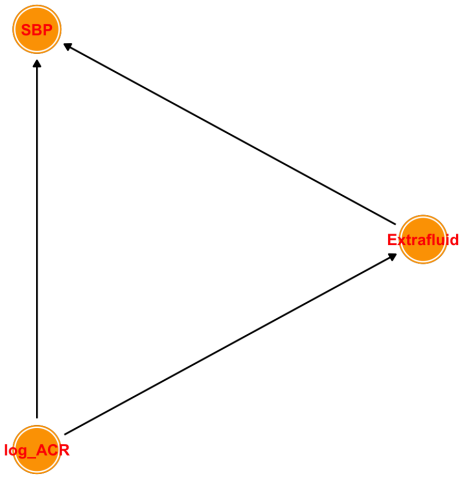
Reduced kidney function measured by elevated serum creatinine levels causes increased blood pressure due to fluid retention.

Rationale**:**

Impaired renal function leads to fluid retention.

Causal Pathway**:**

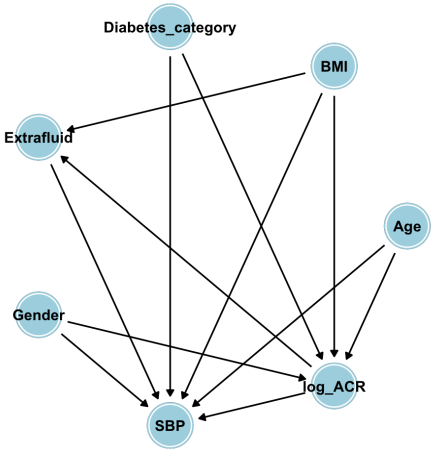
Reduced Kidney Function (Creatinine) → Fluid Retention → Increased Blood Pressure



Causal Assumptions represented by DAG.

Primary outcome

The primary outcome of this study is the association between increased serum creatinine levels and increased systolic. The outcome is measured as Systolic Blood Pressure (SBP) levels as continuous variables and hypertension status as a binary variable, based on clinical cutoffs (SBP ≥ 140 mmHg or DBP ≥ 90 mmHg). The causal effect is estimated by the change in blood pressure per unit increase in serum creatinine.



Secondary outcome

The secondary outcome of this study is to assess additional including quantification of fluid retention as an intermediate outcome between creatinine levels and blood pressure. Differential impacts of reduced kidney function on hypertension across demographic groups, accuracy and predictive power of a model incorporating creatinine levels and additional risk factors to predict hypertension development, and investigating whether pre-existing hypertension leads to elevated creatinine levels will also be assessed.

Exploratory Data Analysis

Data description

The data for this study is obtained from the National Health and Nutrition Examination Survey (NHANES), a publicly available dataset collected by the CDC. The data contains comprehensive health and nutrition data from a nationally representative sample of the US population. Since the BIA (Bioelectrical impedance analysis) data is available only till 2003-04, I took NHANES data also from 2003 to 2004.

Inclusion Criteria:

- Adults aged 20 years and above.
- Availability of albumin and creatinine measurements, blood pressure readings (systolic and diastolic), demographic data, and other relevant covariates like BMI, diabetes status, weight.

Exclusion Criteria:

- Pregnant Women
- Incomplete or missing data on primary variables of interest (creatinine, blood pressure).
- Participants with extreme outlier values that could indicate data entry errors or measurement inaccuracies.

Data variables

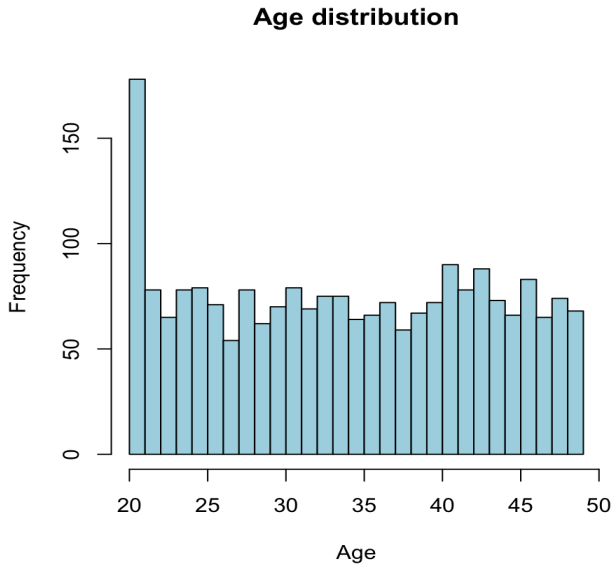
- Primary Exposure Variable (Kidney Function): Creatinine (mg/dL), Albumin (ug/L). Albumin-Creatinine ratio (ACR) is considered as an indicator of kidney health.
- Primary Outcome Variable (Hypertension): Systolic Blood Pressure (SBP) and Diastolic Blood Pressure (DBP) and hypertension status (binary: yes/no). Individuals with SBP $\geq$ 140 mmHg are considered as hypertensive.
- Confounding Variables:
    - Demographic Variables: Age, Sex, weight
    - Health Indicators: BMI, Diabetes Status.

EDA

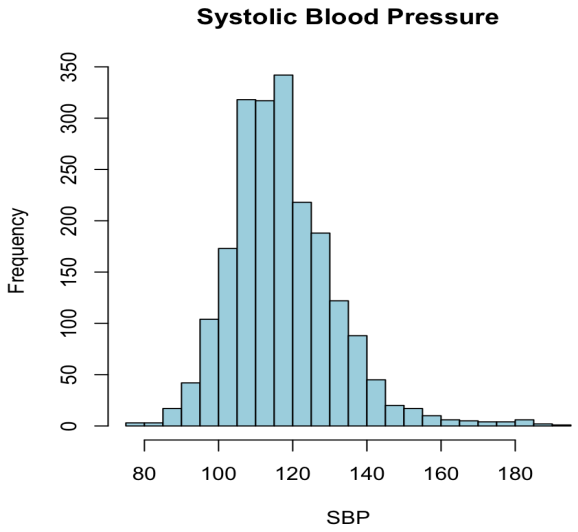
Descriptive Statistics of baseline parameters:

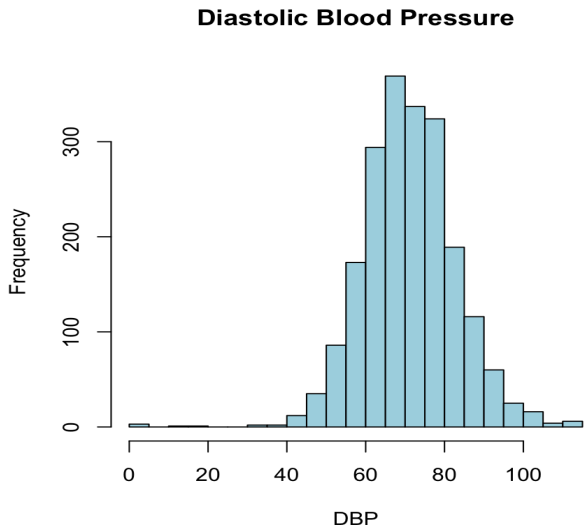
- Continuous Variables: Mean, median, standard deviation (e.g., age, BMI, ACR, SBP, DBP).
- Categorical Variables: Frequencies and percentages (e.g., gender, hypertension prevalence).
- Hypertension status (Yes/No)
- Kidney function categories according to ACR levels.
    - A1: $<$ 30 mg/g - Kidney is functioning normally or with very mild damage.
    - A2: $\geq$ 30 and $\leq$ 300 mg/g - this range indicates early kidney damage.
    - A3: $>$ 300 mg/g - Kidney has significantly high risk of damage.
- Z-test for continuous variables.
- Chi-square test for categorical variables.

Distributions



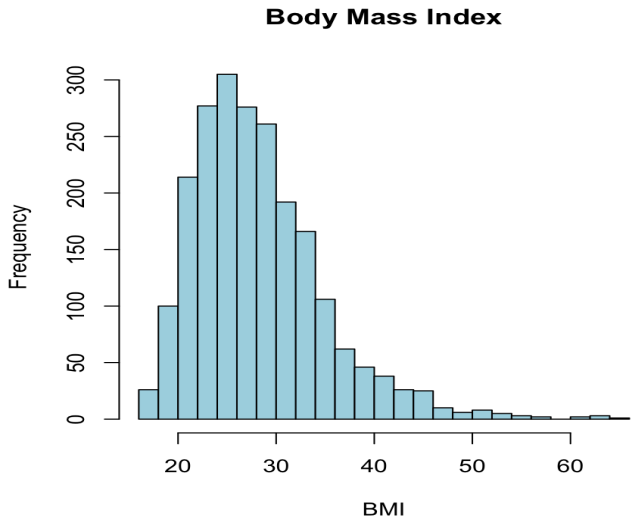
The sample includes adults aged 20 to 50, in which more than 150 individuals are in their early 20s. The rest of the distribution appears relatively uniform, suggesting a balanced age spread within the 20–50 range.





SBP and DBP

SBP appears normally distributed with a mean around 120 mmHg. The right tail includes hypertensive readings (>140 mmHg). DBP also shows a normal distribution, centered around 70–80 mmHg. I can see some outliers present in the data on the left tail.

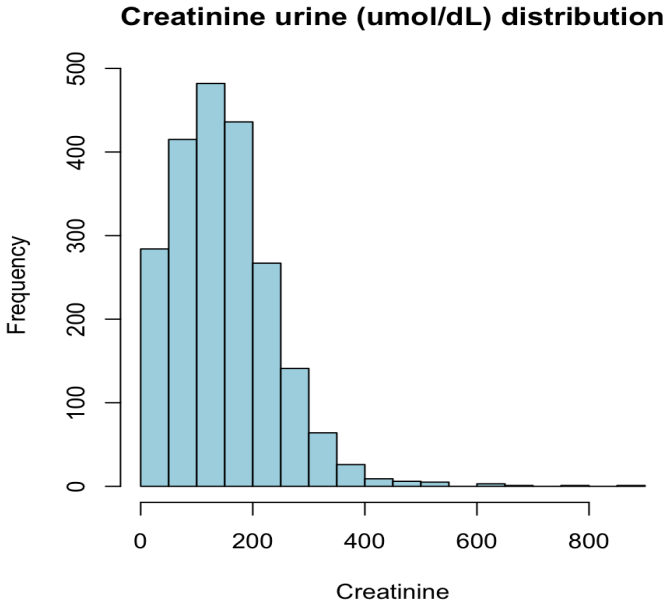


BMI

Most participants in our dataset have a BMI between 25 and 35, suggesting that the majority are either overweight or obese. Very few individuals fall into the underweight category (BMI < 20) or show signs of extreme obesity (BMI > 50), making the distribution right-skewed. Since BMI is closely linked to both high blood pressure and impaired kidney function, it serves as an important factor to include in our analysis.

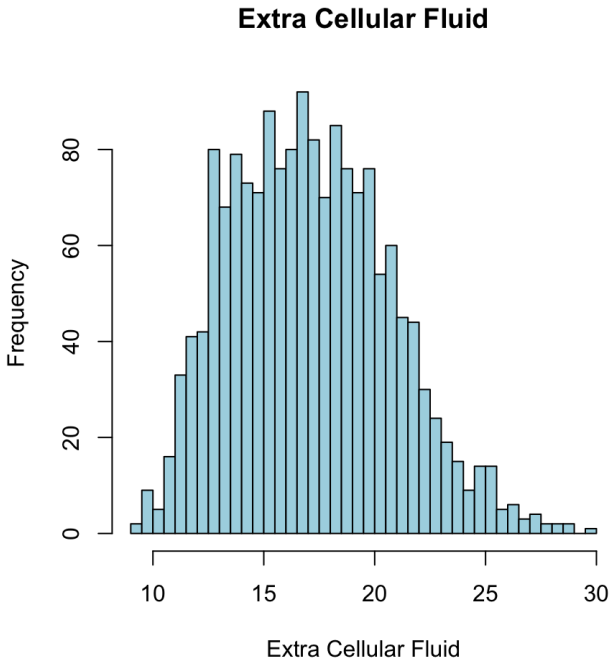
Urine Creatinine level

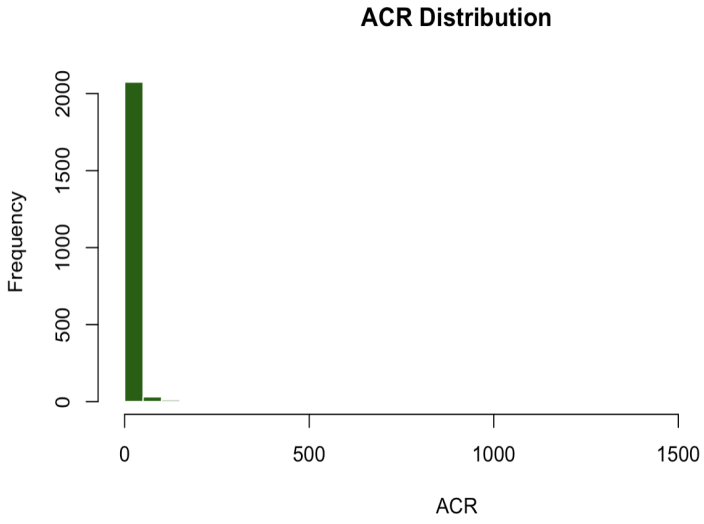
The distribution is right-skewed, with most creatinine levels falling between 50 and 250 µmol/L and peaking around 150 to 200. Higher levels may signal reduced kidney function, particularly when paired with an elevated ACR.

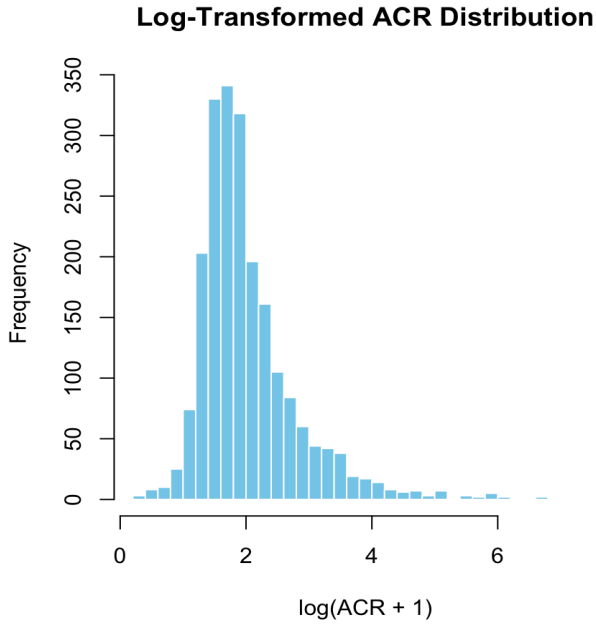


Extra Curricular Fluid

Extracellular fluid (ECF) levels tend to cluster around 15 to 20 liters, though the distribution shows a slight rightward skew. In this analysis, ECF plays a mediating role between kidney function and blood pressure, reflecting how fluid retention often arises when the kidneys aren't filtering efficiently. A few unusually high values above 25 liters may point to cases of fluid overload or potential measurement issues.



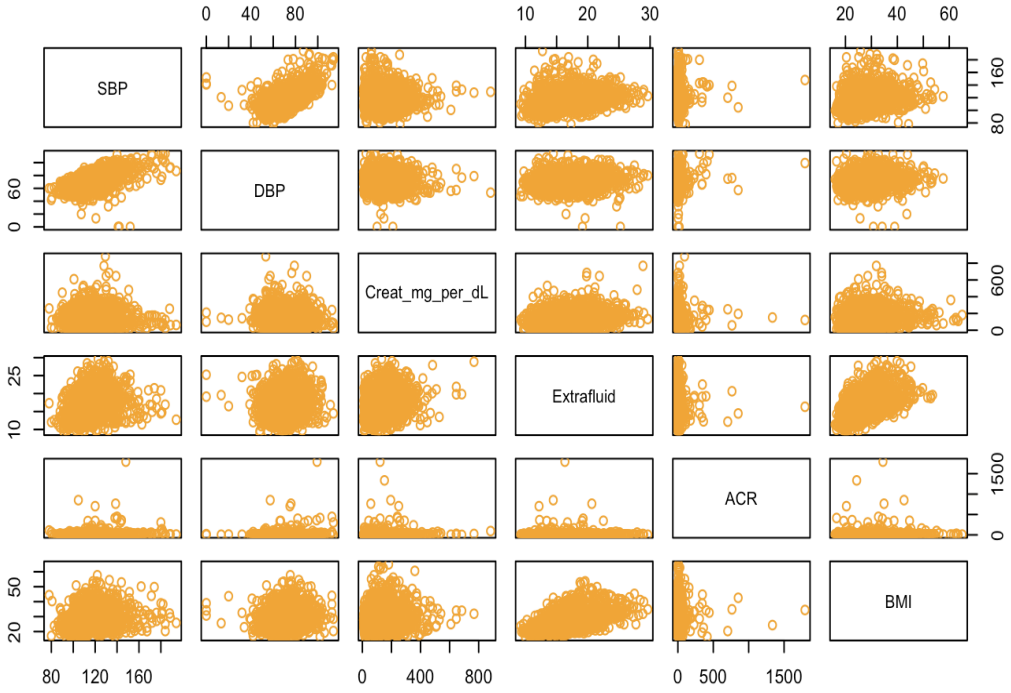




After applying the log transformation, the ACR distribution has become nearly symmetrical, with most values clustering between 1.5 and 2.5. This change effectively reduces the original right skew and pulls in

extreme outliers, helping to stabilize the variance. The transformed distribution is now much better suited for linear regression, correlation studies, and mediation analysis, as it better satisfies assumptions of normality and linearity. It also makes the connection between ACR and outcomes like systolic blood pressure easier to interpret.

Correlations



The scatterplot matrix suggests a clear positive link between systolic blood pressure (SBP) and extracellular fluid, hinting that fluid retention may play a role in raising blood pressure. Urinary creatinine also shows modest positive correlations with both SBP and fluid levels, reinforcing the idea that reduced kidney function can lead to fluid buildup, which in turn elevates blood pressure. In contrast, ACR appears more difficult to interpret in its raw form due to strong right skew and outliers; this makes a log transformation or categorizing it into clinical stages (A1–A3) more appropriate for analysis. BMI shows a slight positive association with SBP, which is consistent with the established link between obesity and hypertension. Altogether, the visual patterns point toward a possible mediating effect of fluid retention in the relationship between kidney health and blood pressure.

Correlation matrix

	SBP	DBP	Creat_mg_per_dL	Extrafluid	log_ACR	BMI	Weight
SBP	1.00000000	0.542570824	0.022630486	0.2736795	0.12679120	0.25877058	0.28806748
DBP	0.54257082	1.000000000	-0.006958982	0.1388562	0.10820461	0.17242800	0.18909604
Creat_mg_per_dL	0.02263049	-0.006958982	1.000000000	0.2168657	-0.03380942	0.10621966	0.19029093
Extrafluid	0.27367953	0.138856194	0.216865738	1.0000000	-0.14921208	0.50110112	0.83090406
log_ACR	0.12679120	0.108204612	-0.033809420	-0.1492121	1.00000000	0.06052194	-0.03523329
BMI	0.25877058	0.172428003	0.106219664	0.5011011	0.06052194	1.00000000	0.85960281
Weight	0.28806748	0.189096044	0.190290927	0.8309041	-0.03523329	0.85960281	1.00000000

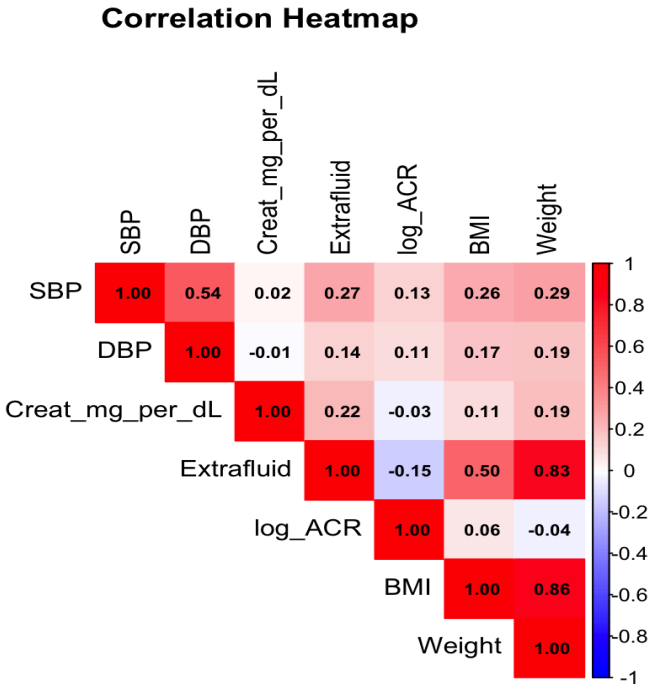
SBP shows a modest correlation with extracellular fluid (r = 0.27), lending some support to the hypothesis that fluid retention may contribute to higher blood pressure, though the effect isn't particularly strong. Its correlation with log-transformed ACR is even weaker (r = 0.13), suggesting only a mild association between kidney damage and blood pressure when considered alone. SBP is also moderately correlated with BMI (r = 0.26) and weight (r = 0.29), which aligns with well-established links between obesity and hypertension.

Creatinine shows a positive correlation with extracellular fluid (r = 0.22), supporting the biological link between reduced kidney function and fluid buildup. However, its correlation with SBP (r = 0.02) and DBP (r = -0.01) is minimal, suggesting that any direct effect on blood pressure is likely weak. This reinforces the need to explore mediated pathways or potential non-linear effects. As expected, creatinine has only weak associations with BMI and weight, since it tends to reflect muscle mass and hydration status more than body fat.

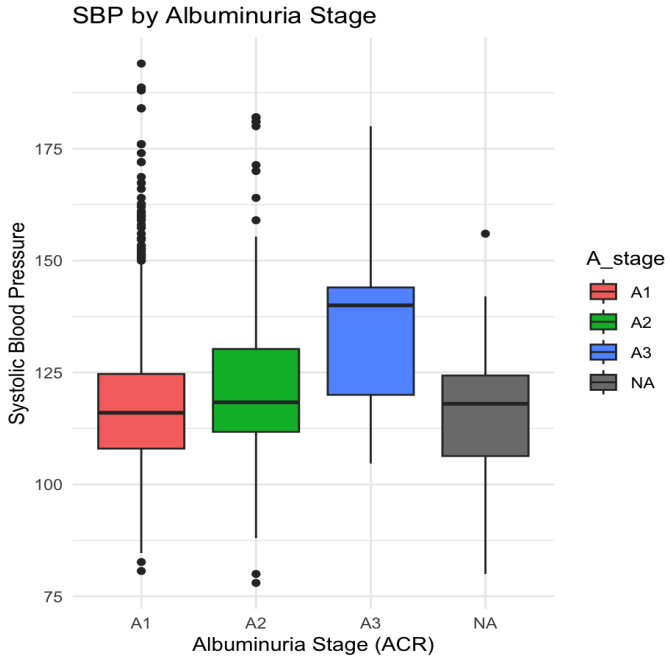
Fluid retention shows a strong correlation with weight (r = 0.83) and a moderate one with BMI (r = 0.50), which makes sense given that larger individuals tend to retain more fluid. Its positive association with SBP (r = 0.27) also supports the idea that fluid buildup may act as a mediator between body size or kidney function and elevated blood pressure.

Log_ACR shows weak correlations with SBP (r = 0.13) and DBP (r = 0.11), and a slight negative correlation with extracellular fluid (r = -0.15), which fits the idea that ACR reflects kidney damage rather than fluid overload. It's essentially uncorrelated with weight or BMI.

We can say that log_ACR and Creatinine are not strongly directly correlated with blood pressure but may act via Extrafluid supports our mediation hypothesis.



The box plot below shows a clear upward trend in median SBP across albuminuria stages, rising from A1 (normal) to A3 (severely elevated ACR). Participants in the A3 group not only have higher median SBP but also greater variability, suggesting a strong link between advanced kidney damage and elevated blood pressure. A1 and A2 groups have lower SBP levels, with A1 representing the healthiest profile. Interestingly, more outliers appear in A1, which may point to undiagnosed hypertension or unrelated risk factors.The plot also supports our hypothesis that worsening kidney function indicated by ACR is associated with higher blood pressure.



Kidney Damage Stage	Creatinine(mean)	ECF(mean)	SBP(mean)
A1	148	17.1	117
A2	157	16.7	124
A3	166	16.5	136

Group means show that creatinine levels rise from A1 to A3, indicating declining kidney function. SBP also increases notably across these stages, providing strong evidence of a link between kidney damage and elevated blood pressure. Interestingly, extracellular fluid shows a slight decrease, which may seem unexpected. To better understand its role as a mediator in the kidney function–SBP relationship, we're conducting a mediation analysis in the next section.

Statistical Modeling, Causal Analysis and Results

Models

Statistical Models:

- Multivariable Linear Regression: To estimate the association between creatinine levels and continuous blood pressure outcomes using beta coefficients and p-values. This is used to quantify direct association between Creatinine/log_ACR and SBP/DBP, adjusting for confounders.

Variable	Estimate	p-value	Result
Creat_mg_per_dL	-0.0026	0.440	There is no meaningful linear relationship between Creatinine and SBP.
Age	0.405	<2e-16	A high significance is seen where SBP increases by ~0.41 mmHg per year of age.
Gender	-6.00	<2e-16	Females have ~6 mmHg lower SBP than males on average.
BMI	0.533	<2e-16	Each unit increase in BMI is associated with ~0.53 mmHg higher SBP.
Diabetes	0.649	0.716	There is no significance.

The results show that age, gender, and BMI have a clear and significant impact on systolic blood pressure. SBP tends to rise with age and higher BMI, while women generally have lower SBP than men, about 6 mmHg lower on average. Creatinine and diabetes don't show meaningful effects in this model, which supports the earlier idea that creatinine may influence SBP indirectly rather than directly.

For log_ACR

Predictor	Coefficient	p-value	Result
log_ACR	3.36	<2e-16	Each unit increase in log_ACR is associated with a 3.36 mmHg increase in SBP.
Age	0.41	<2e-16	Each additional year of age increases SBP by 0.41 mmHg.
Gender	-6.91	<2e-16	Females have ~6.9 mmHg lower SBP than males on average.
BMI	0.51	<2e-16	Each unit increase in BMI is associated with ~0.51 mmHg higher SBP.
Diabetes	-1.19	0.501	There is no significance.

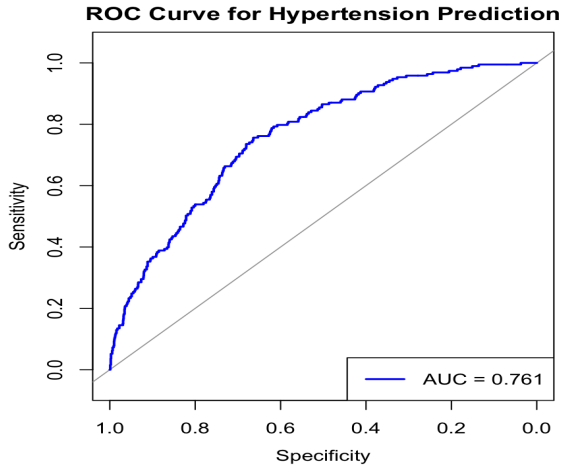
The regression model shows that log-transformed ACR is a strong and significant predictor of systolic blood pressure, even after accounting for age, gender, BMI, and diabetes. For every 1-unit increase in log_ACR, SBP rises by about 3.36 mmHg ($p < 0.001$). This model explains nearly 20% of the variation in SBP, performing better than models that include creatinine alone. Age, gender, and BMI remain important factors, while diabetes doesn't appear to have a meaningful effect in this case.

- Logistic Regression: The outcome will be hypertension status. To estimate the odds ratio, confidence intervals, and p-values of hypertension based on creatinine levels.

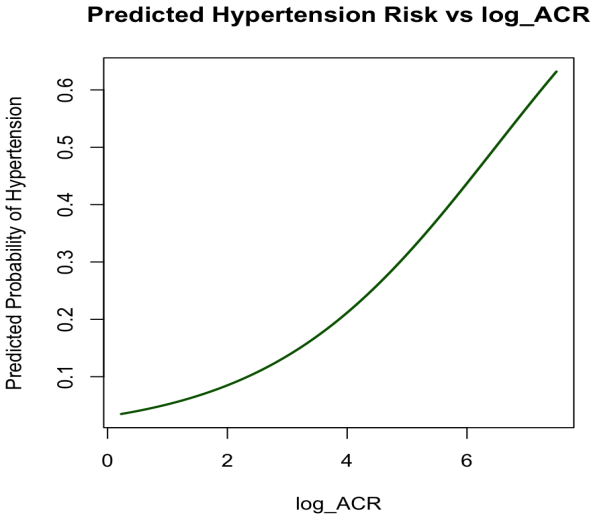
Variable	Estimate	exp(β)	p-value	Result
log_ACR	0.531	1.70	2.4e-10	Each unit increase in log_ACR increases the odds of hypertension by 70%.
Age	0.089	1.093	<2e-16	Each additional year of age increases odds by 9.3%.
Gender	-0.486	0.615	0.003	Females have 38.5% lower odds of hypertension than males.
BMI	0.060	1.062	9.86e-07	Each unit increase in BMI increases hypertension odds by 6.2%.
Diabetes	-0.403	0.668	0.30	Not statistically significant.

The logistic regression results show that log-transformed ACR is a strong and significant predictor of hypertension. Each 1-unit increase in log_ACR raises the odds of having hypertension by about 70% ($p < 0.001$). Age and BMI also increase the risk, while females are significantly less likely than males to be hypertensive. Diabetes didn't show a meaningful effect in the model. The results highlight ACR as an important, independent marker of hypertension risk.

ROC Curve



In medical prediction, an AUC between 0.75 and 0.8 is generally considered good. The above ROC curve for our logistic regression model, using log-transformed ACR, age, gender, BMI, and diabetes shows an AUC of 0.761. This suggests the model does a solid job distinguishing between people with and without hypertension, demonstrating good overall predictive performance.



The figure illustrates how the predicted risk of hypertension increases with higher log_ACR levels. People with low log_ACR values (around 0–2) have less than a 20% chance of being hypertensive, while those with values near 6 face a risk over 60%. The curve is not linear, risk rises more sharply at higher ACR levels showing that even small increases in ACR can have a big impact. This highlights the importance of early ACR monitoring to catch and manage hypertension risk before it escalates.

Sensitivity Analysis:

Stratified logistic regressions by gender showed that the link between log_ACR and hypertension held in both groups, with a slightly stronger effect observed in females. Including or excluding BMI in the model made little difference to the overall findings.

Model Validation:

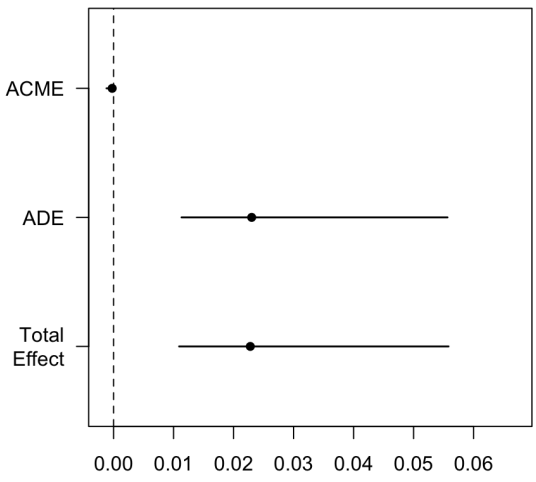
- Perform Cross-Validation to evaluate model performance and generalizability.

Metric	Value	Result
ROC AUC	0.753	The model correctly distinguishes hypertensive vs non-hypertensive 75% of the time.
Sensitivity	0.996	Very high sensitivity of the model correctly identifies almost all hypertensive individuals.
Specificity	0.036	Specificity is very low which means that the model struggles to correctly identify non-hypertensive individuals (high false positives).

Using 5-fold cross-validation, the model achieved an AUC of 0.753, reflecting good overall performance. It showed extremely high sensitivity (99.6%), meaning it correctly identified nearly all individuals with hypertension. However, its specificity was low (3.6%), indicating that it also flagged many non-hypertensive individuals as at risk suggesting the model tends to over-predict hypertension (more false positives).

Mediation Analysis

Effect Type	Estimate	95% CI Lower	95% CI Upper	p-value	Significance
ACME (Indirect)	-0.00023	-0.00117	0.00000	0.2965	Not Significant
ADE (Direct)	0.02302	0.01133	0.06000	0.0050	Yes
Total Effect	0.02279	0.01093	0.06000	0.0038	Yes
Proportion Mediated	-0.01007	-0.07448	0.01000	0.3003	Not Significant



We ran a causal mediation analysis to see if fluid retention helps explain the link between ACR and systolic blood pressure (SBP). The results showed a significant direct effect. Higher ACR was clearly associated with higher SBP ($p = 0.005$). The total effect was also significant ($p = 0.0038$), confirming a strong overall relationship. However, the indirect effect through fluid retention wasn't significant ($p = 0.296$), and the proportion mediated was minimal and even slightly negative. This suggests that fluid retention doesn't

meaningfully explain the ACR–SBP connection. Instead, ACR seems to affect blood pressure more directly.
The same mediation analysis we did by taking Creatinine and log_ACR also but the results were similar to the ACR

Hypothesis Tests, Correlations and Results

1. Gender vs Hypertension (Chi-Square Test)

H_0: There is a relationship between gender and hypertension status.

A chi-square test was conducted to examine the relationship between gender and hypertension status. The test was not statistically significant ($p = 0.292 > 0.005$), indicating that there is no evidence of a significant association between gender and hypertension in this dataset.

2. SBP Across ACR Stages (ANOVA and Tukey HSD)

H_0: There is a significant difference in mean SBP across stages.
A one-way ANOVA showed a highly significant difference in mean systolic blood pressure across ACR stages ($p < 0.0001$).

Post-hoc Tukey HSD tests revealed:

- Mean SBP in A2 was 6.41 mmHg higher than A1 ($p = 0.00003$),
- Mean SBP in A3 was 19.32 mmHg higher than A1 ($p = 0.000004$),
- Mean SBP in A3 was 12.91 mmHg higher than A2 ($p = 0.0064$).

This supports that worsening kidney damage (higher ACR stages) is associated with significantly higher SBP.

3. SBP in Diabetics vs Non-Diabetics (T-Test)

H_0: There is a significant difference in mean SBP between diabetics and non-diabetics.

An independent t-test showed a statistically significant difference in mean SBP between diabetics and non-diabetics ($p = 0.0027$), $p < 0.005$.

- Non-diabetics had a mean SBP of 117.30 mmHg,
- Diabetics had a mean SBP of 123.56 mmHg.

Thus, diabetics had significantly higher systolic blood pressure than non-diabetics.

4. BMI by Hypertension Status (T-Test)

H_0: There is a significant difference in BMI between hypertensive and non-hypertensive individuals.

An independent t-test revealed a highly significant difference in BMI between hypertensive and non-hypertensive individuals ($p < 0.000001$). So, accepting the null hypothesis.

- Normotensive individuals had a mean BMI of 27.71,
- Hypertensive individuals had a mean BMI of 30.70.

Thus, individuals with hypertension had significantly higher BMI compared to those without hypertension.

5. SBP Across Age Groups (ANOVA and Tukey HSD)

A one-way ANOVA showed a highly significant difference in SBP across age groups ($F = 72.21$, $p < 0.0001$).
Post-hoc Tukey HSD comparisons indicated:

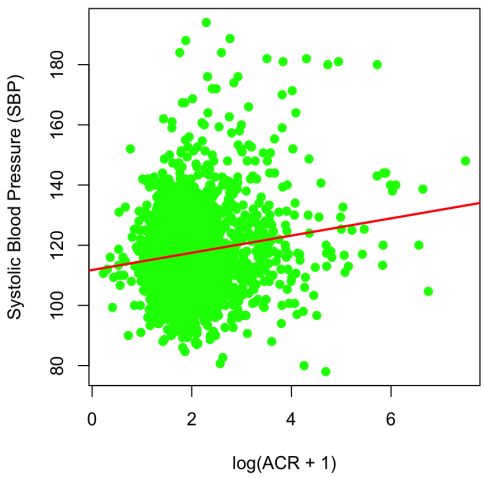
- Mean SBP in the 30–39 group was 2.56 mmHg higher than 20–29 ($p = 0.0024$),
- Mean SBP in the 40–49 group was 8.74 mmHg higher than 20–29 ($p < 0.0001$),
- Mean SBP in the 40–49 group was 6.18 mmHg higher than 30–39 ($p < 0.0001$).

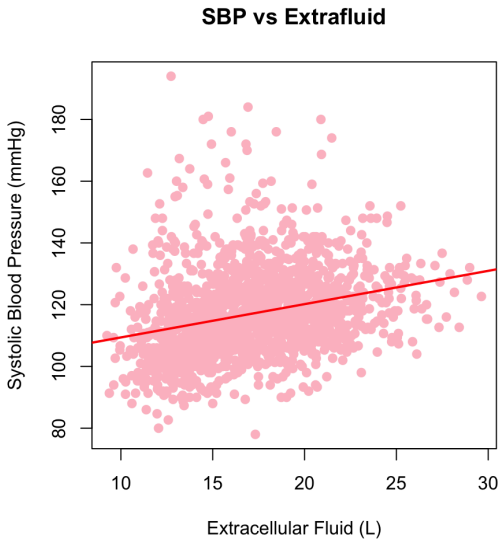
Thus, older age groups had significantly higher systolic blood pressure, confirming that SBP increases with age.

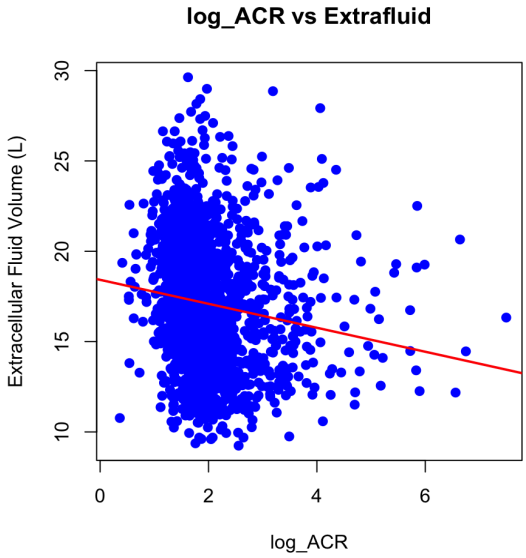
Relationship	Correlation	p-value	Strength	Result
log_ACR vs Extrafluid	-0.145	3.12e-09	Weak Negative	As log_ACR increases, Extrafluid slightly decreases.
Extrafluid vs SBP	0.274	<2.2e-16	Moderate Positive	Higher Extrafluid is associated with higher SBP
log_ACR vs SBP	0.156	1.69e-12	Weak Positive	Higher log_ACR is associated with higher SBP

Correlation between log_ACR and Extracellular Fluid, log_ACR and SBP, Extrafluid and SBP

Correlation Analysis: Correlation coefficient analysis to assess the relationship between serum creatinine and blood pressure.







Pearson correlation analysis showed a weak but statistically significant negative relationship between log_ACR and extracellular fluid ($r = -0.145$, $p < 0.001$), suggesting that more severe kidney damage doesn't always mean more fluid retention. On the other hand, extracellular fluid was moderately correlated with systolic blood pressure ($r = 0.274$, $p < 0.001$), pointing to a clear link between fluid buildup and higher BP. Log_ACR also had a weak yet significant positive correlation with SBP ($r = 0.156$, $p < 0.001$), supporting the idea that kidney dysfunction contributes directly to elevated blood pressure.

Software

Statistical models like linear regression and logistic regression and R libraries like NHANES, nhanesA, haven, tidyverse, ggplot2, doBy, tidyr, caret, corrplot, mediation, pROC, and dplyr.

Conclusions

Our analysis highlights a clear link between kidney dysfunction and elevated blood pressure. Elevated albumin-to-creatinine ratio (ACR), particularly when log-transformed, emerged as a strong and

independent predictor of systolic blood pressure and hypertension risk, even after accounting for key factors like age, gender, BMI, and diabetes status. Interestingly, while ACR showed a direct and significant impact on blood pressure, our mediation analysis found that this effect is not meaningfully explained by fluid retention. So, our causal relationship is only partially supported. This suggests that albuminuria influences blood pressure through mechanisms beyond simple volume overload perhaps through vascular or inflammatory pathways.

Extracellular fluid was moderately associated with blood pressure and strongly tied to body weight and BMI, supporting its role in hypertension risk but not as a central mediator in the ACR–SBP relationship. Creatinine, on the other hand, showed minimal direct association with blood pressure, reinforcing the idea that it may reflect underlying kidney function without directly influencing SBP.

The predictive models, including logistic regression and ROC analysis, demonstrated good performance overall, with AUC values around 0.75. While sensitivity was high, effectively identifying those at risk, specificity was low, highlighting the trade-off between capturing all hypertensive individuals and minimizing false positives.

Altogether, my findings emphasize the clinical importance of ACR as a sensitive marker of hypertension risk, independent of fluid retention. They also underscore the need for early detection and targeted interventions that address kidney health in managing blood pressure and preventing cardiovascular complications.

References

1. Centers for Disease Control and Prevention (CDC). (n.d.). National Health and Nutrition Examination Survey (NHANES).
2. Ma, X., Lin, F., Liu, Y., Lu, L., Wu, Y., Chen, Y., & Feng, Y. (2023). Serum Creatinine and Cardiovascular Risk in Hypertensive Patients. *Frontiers in Cardiovascular Medicine, 10*, 1162480.
3. Shi, W., Cao, J., & Li, Z. (2023). Urinary Albumin-to-Creatinine Ratio and Hypertension Prevalence. J*ournal of Clinical Hypertension, 25*(8), 870–877.
4. Qian, Y., Wang, Y., Wang, X., Wang, X., & Wang, Y. (2023). Serum Creatinine/Cystatin C Ratio and Cardiovascular Disease. *Reviews in Cardiovascular Medicine, 24*(9), 275.
5. Zhao, J., Booth, J. N., Sims, M., Muntner, P., & Shimbo, D. (2024). Blood Pressure Control in Chronic Kidney Disease. *Journal of the American Heart Association, 13*(5), e034568.
6. https://pubmed-ncbi-nlm-nih-gov.ezproxy.yu.edu/35652832/ for BIA data.
7. OpenIntro Statistics. Fourth Edition. David Diez, Randel, M., C., Barr C., D.