Uncovering hidden dynamics and links between Heart disease and Chronic kidney disease through statistical analysis and machine learning tools

Itamar Zernitsky - 328437702
Yaniv Ilan - 215971441

July 2025

Abstract

In this article, we aim to examine the existence of hidden relationships and dynamics for heart disease and chronic kidney disease. In addition, our central hypothesis in the article is not only that such relationships and dynamics exist, but that the relationship is unidirectional. We hypothesize that patients with a heart disease will probably also develop chronic kidney disease, but the reverse is not true. We work with two databases during the study, the UCI heart disease dataset and the UCI chronic kidney disease dataset, on which we perform statistical research using statistical tests, linear and nonlinear regressions, as well as logistic regression models in order to create a common data set we can work on. We were indeed able to diagnose characteristics of both diseases and train models that make predictions with high accuracy, and, with their help, produce common and global data for both diseases. The results we reached in the article indicate that there is indeed a significant relationship between a heart disease and chronic kidney disease, and that the relationship is indeed clearly unidirectional. A heart disease causes the development of chronic kidney disease, and that there are indeed dynamics between the two diseases. The code for this project is available at: `https://github.com/Itamarzer/Statistical-Theory.git`

Keywords— heart disease, ckd - Chronic kidney disease, Machine learning, regression, statistical tests, population research.

1 Introduction

Heart disease and chronic kidney disease have been the subject of research for some time. Many studies have been conducted on these diseases separately to understand their characteristics and the factors associated with their development and progression. Other studies have not only investigated the diseases separately but have also explored the connections between these diseases (see Chew et al). In this study, we aim to refine our understanding of the connections between these diseases, and therefore, we first need to conduct additional research on each disease separately. To better understand the two diseases, we will conduct population studies on them, that is, understanding and studying the different groups in which each disease is prevalent, as well as the populations in which we can see and deduce from the dynamics of those hidden connections between the diseases. The databases we used in this work are the UCI heart disease dataset and the UCI Chronic kidney disease dataset, and, we also based part of our work on the results of a study conducted on the Nhanes dataset (National Health and Nutrition Examination Survey). Since these two datasets contain information on only one of the diseases, we also used machine learning methods and models, as well as linear and nonlinear regression, and logistic regression. In this study we have to goals. First, we aim to find certain population groups that will help us understand and explain the hidden dynamics between the diseases and of each disease separately. We will do this through statistical research and by applying different statistical tests between the populations. Second, to refine the understanding between these two diseases. our central hypothesis, which can be divided into two parts, is: A. There is indeed a relationship between the two diseases, and this relationship is statistically significant. B. This relationship is, in most cases, one-sided; that is, we hypothesize that people with heart disease will probably also have chronic kidney disease as a result of their heart disease, but chronic kidney disease patients will not necessarily develop heart disease as a result.

2 Methods

2.1 Basic initial research

First, we performed a normality test for the two different databases and their characteristics using the Shapiro-Wilks test and a Q-Q plot. We then performed Person correlation, Spearmann correlation, and Cramer's V correlation for the categorical characteristics.

2.2 The UCI Heart disease dataset analysis

First of all, we wanted to address several characteristics of the dataset to refine our analysis further, so we focused on two specific characteristics: first, the severity of atypical chest pain, which is described as options: typical angina, atypical angina, non-anginal, asymptomatic by survey respondents, which we coded on a scale of 1-4, respectively. Second, the severity of heart disease, which is described on a scale of 0-3. The treatment of these characteristics was done to check whether we can refer to variants of a a heart disease in the database in a binary form of having a heart disease or not having a heart disease, rather than in the way in which they are represented as not having heart disease and certain degrees of heart disease. We ran a chi-square test for these two characteristics and their characteristic groups to test for independence and a Levene's test to test for equality of variance. For the characteristic of chest pain type, we tested whether older patients tend to experience atypical chest pain (pain types 2-4) in comparison to younger patients, since the age characteristic is used later to predict diseases and has a significant correlation with it. In order to test this, we ran the parametric t-test and the non-parametric: Mann -Whitney U test and the permutation test, which do not assume equal variance. For the characteristic of the heart disease severity, we tested whether the average amount of cholesterol in serum also has a significant correlation with the heart disease, and the severity of the heart disease. We did it using the parametric One-way ANOVA test and the non-parametric Kruskal Wallis test. From the results of this section, which will be presented below, we decided to represent the heart disease characteristic in a binary way - with/without heart disease.

We performed Levene's tests on the disease variable (num) to assess the homogeneity of variance. For categorical variables, chi-square tests of independence were used to assess associations with the presence of heart a disease. To examine the common association between heart disease and several characteristics (trestbps, chol, thalch, oldpeak), we conducted a multivariate analysis of variance (MANOVA). To control for potential confounding variables (age, sex, cp, fbs), a multivariate analysis of covariance (MANCOVA) was also conducted. For each dependent variable separately, analysis of covariance (ANCOVA) was used to estimate the effect of disease while adjusting for those covariates. Finally, we performed subgroup and population-level analysis. That is, we conducted a comprehensive population subgroup analysis by defining clinically meaningful groups (e.g., "older men," "high cholesterol," "asymptomatic," etc.). For each group, we calculated the prevalence of heart disease and the mean severity of disease. We performed chi-square tests between each group and the presence of disease, and we visualized the results using bar graphs and box plots. To adjust for multiple testing, p-values In this part were corrected using Bonferroni, Holm, and Benjamini–Hochberg (FDR) procedures.

2.3 The UCI Chronic kidney disease dataset analysis

In order to conduct a future study of the two populations, we tested two hypotheses for the CKD database. The first hypothesis is that CKD patients have a lower hemoglobin level than the average healthy person, which is 12.0 g/dL for women and 13.5 g/dL for men. We took a weighting of both - 12.75 g/dL. Second, we hypothesize that the amount of red blood cells in the blood for CKD patients is lower than the average amount in a healthy person, ranging from 4.2 to 6.1 million cells per microliter (mcL). For both hypotheses, we first performed Levene's test and chi-square test to check for equality of variances and independence of the respective populations. We then performed a two-tailed t-test to check whether there is indeed a significant difference between the means and then a one-tailed t-test. We then ran a one-tailed and a two-tailed Mann Whitney U tests to not rely on a normal distribution of the samples. Finally a permutation test was performed in order to not rely on the assumption of equality of variances.

For population research, we started with Group Comparisons. We assessed differences in the continuous variables between CKD and non-CKD groups using the Mann-Whitney U test due to potential non-normality. We calculated effect sizes as Cohen's d, with values interpreted as small (0.2–0.5), medium (0.5–0.8), or large (greater than 0.8). For the categorical variables we evaluated associations with the CKD status using Chi-square tests of independence. We then adjusted the comparisons using Bonferroni, Holm, and Benjamini-Hochberg (FDR) methods. We conducted a Multivariate analysis of variance (MANOVA) to test differences in multiple dependent variables (Serum Creatinine,

Hemoglobin, Blood Urea, Blood Glucose Random) across the CKD status. In addition, we performed a Multivariate Analysis of Covariance (MANCOVA) to control for covariates (Age, Blood Pressure, Hypertension, Diabetes Mellitus). In both analyses, we used a formula-based approach to handle column names containing spaces, and the p-values were extracted and corrected for multiple comparisons. We then conducted Univariate analysis of covariance (ANCOVA) for each dependent variable, adjusting for the same covariates. The P-values were corrected using Bonferroni, Holm, and FDR methods. We also defined subgroups based on clinical thresholds (e.g., Age greater than 50, Blood Pressure greater than 140, Serum Creatinine greater than 1.2) and categorical conditions (e.g., Hypertension, Diabetes Mellitus). We calculated for each subgroup the prevalence of CKD, mean Serum Creatinine, and mean Hemoglobin and applied chi-square tests to assess associations between a subgroup membership and the CKD status, with p-values corrected for multiple comparisons.

2.4 Cross data analysis and Hypothesis validation

First, we performed learning on the CKD dataset by dividing it into training data and test data to predict whether the patient has/does not have CKD using the parameters: 'Hemoglobin', 'Packed Cell Volume', 'Red Blood Cell Count', 'Hypertension' and 'Diabetes Mellitus' because they showed high and significant correlations (person/spearmann) with the CKD feature. This training was performed by comparing the performance of the models: logistic regression, KNN, SVM, Random forests, Xgboost, and light gbm using optuna for hyperparameter optimization and selecting the model that maximizes the F1 score result, while checking that no overfitting was performed. Since we have two databases, one containing information for those with/without heart disease and the other for those with/without chronic kidney disease, but not a database containing both, and since they do not have the same characteristics that we used to perform learning on the ckd dataset, we relied on hemoglobin and the amount of red blood cells in the study on the NHAMES database (see Walker et.al). We subtracted from the hemoglobin a normal noise with a mean of 0 and a variance of 5^2 if the person has a heart disease, and for the amount of red blood cells a normal noise with a mean of 0 and a variance of 0.1.

$$\text{Hemoglobin} = 13.5 + 1.5 \cdot \text{sex} - 0.007 \cdot \text{age} - \epsilon_H \, \mathbf{1}_{\{\text{HD}\}}$$

and

$$\text{RBC} = 0.45 + 0.23 \cdot \text{Hemoglobin} + 0.35 \cdot \text{sex} - 0.005 \cdot \text{age} - \epsilon_R \, \mathbf{1}_{\{\text{HD}\}}$$

where $\mathbf{1}_{\{\text{HD}\}}$ equals 1 if the patient has a heart disease and 0 otherwise; $\epsilon_H \sim \mathcal{N}(0, 5^2)$ and $\epsilon_R \sim \mathcal{N}(0, 0.1^2)$.
That is, based on the initial statistical study we performed on the ckd dataset, assuming that heart disease also has similar characteristics, but using the normal distribution (which can give negative values for the noise we chose) in case this assumption is incorrect. For Hypertension, we used a logistic regression with the parameters age, hemoglobin, and red blood cells count. The same without red blood cell count for Diabetes Mellitus. For the Packed cell volume, we used a linear regression with the parameters age, hemoglobin, and Red blood cell, then applied optimization on the degree of the regression and used non - linear regression.
Finally, we used the logistic regression and the linear or non linear regression to add the features to the heart disease dataset. Then, the model that was trained on the ckd dataset was used to make predictions on the heart disease dataset. Thus, obtaining a single database containing information about patients related to both diseases - on which we used z-tests and the non-parametric binomial test to test the ratio of patients with a heart disease but not kidney disease and vice versa. This was done to test our research hypothesis, as well as studying populations for patients with only one of the diseases and patients with both.

3 Results

3.1 correlation and Normality

For both databases, both for heart disease and chronic kidney disease, according to the q-q plots and for the Shapiro-Wilks tests, we obtained that all features are not normally distributed, and therefore we will have to back up every parametric test in the project with a non-parametric test. As for the correlations, also for both databases, all correlations, which are Pearson, Spearman, and Cramer's V, gave significance for the feature of the mother being a carrier/non-carrier of the disease (see Figure 2).

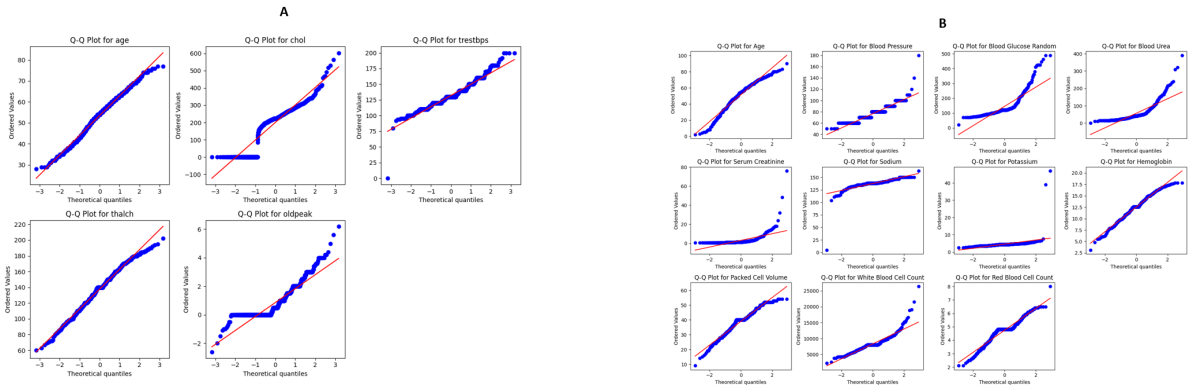

Figure 1: A.Q-Q plots for the heart disease dataset numerical features. B. Q-Q plots for the CKD dataset numerical features.

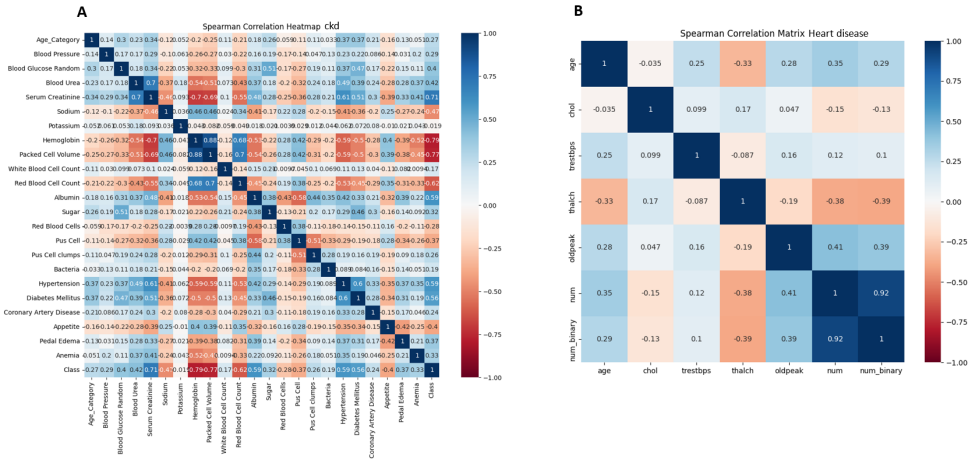

Figure 2: A. Heatmap for Spearman correlation of Chronic Kidney disease dataset features. B. Heatmap for Spearman correlation of Heart disease dataset features.

3.2 Statistical tests analysis

First, in examining our two hypotheses for the Heart disease dataset, the first that older patients are more likely to experience atypical chest pain (type 2, 3 or 4) compared to younger patients who experience typical angina (type 1) and the second that there is no significant difference between mean serum cholesterol levels and the severity of heart disease. For the first hypothesis, we obtained three tests: the t-test, the Mann-Whitney U test (non-parametric), and the Permutation test. In all three, we got P-values, which are greater than 0.05 ($6.03e-2$, $9.95e-1$, and $9.85e-1$ respectively), meaning that there is no significant difference in age between people who experience typical versus atypical pain, so we could transfer it to binary form. For the second hypothesis, we obtained three tests: one-way ANOVA, Kruskal-Wallis, and the permutation test. Once again, in all three, we got P-values, which are greater than 0.05 ($3.39e-1$, $1.66e-1$, and $3.31e-1$ respectively), meaning that average cholesterol levels do not differ significantly between severity levels, so we could transfer it to binary form. Second, in examining our two hypotheses for the Chronic Kidney Disease dataset, is that CKD patients have hemoglobin levels significantly lower than the normal population standard (12.0 g/dL for women, 13.5 g/dL for men; we took a weighted average of both - 12.75). Also, the amount of red blood cells for CKD patients is lower than the average amount for a healthy person (4.2 to 6.1 million cells per microliter (cells/mcL), We took the average of the ranges. For both hypotheses, we ran the t-test, Mann-Whitney U test for the second hypothesis, and Wilcoxon signed rank test for the first one, and Permutation test, and for the number of red blood cells, we tested using the Kolmogorov-Smirnov test and the Anderson-Darling test to see if the distribution for those with/without CKD was the same. For the first hypothesis, we obtained significant results, that is, all $P-values < 0.05$, which are $1.28e-31$, $3.53e-26$, and $p-value < 1e-50$

respectively. For the second hypothesis, we obtained significant results, that is, all $P - values < 0.05$, which are $2.9e-36$, $4.5e-35$, and $p-value < 1e-50$ respectively. These results will be used later when we perform regressions and add features to the heart disease dataset that correspond to those in the CKD dataset based on the work done on the NHAMES database (see Guralnik et. al).

Now, for the heart disease dataset, we found in the ANCOVA analysis that each clinical outcome (trestbps, chol, thalch, oldpeak) is significantly associated with heart disease. And also for the MANCOVA analysis, that after adjusting for age, sex, cp, and fbs, the effect of num, whether the patient has heart disease or not, on the dependent variables remains significant. This means that the group differences across num categories cannot be explained solely by differences in age, sex, chest pain type, or fasting blood sugar. So, the num variable has a strong and independent multivariate effect on the outcomes, beyond what can be attributed to demographic or clinical covariates. Across multiple demographic and clinical subgroups (e.g., Old Males, High FBS, Asymptomatic CP, etc.), heart disease prevalence is consistently significantly associated. For example, old males have 80.8% prevalence and Old Males with High Cholesterol have 81.9% prevalence. All Chi-square tests return $p < 0.0001$, suggesting a strong statistical association. In contrast, old Females have the lowest prevalence (0.211), suggesting a lower risk or different disease presentation. For the CKD dataset population analysis, older patients have a high CKD prevalence (73.8%), with elevated Serum Creatinine and reduced Hemoglobin, consistent with CKD pathophysiology. The Chi-square test, which has yielded a significant result, meaning $p-value < 1e-6$ indicates a strong association between age, (greater than 50) and CKD, similar to high-risk groups in the heart disease analysis (e.g., Old Males).

3.3 Learning hidden dynamics between the diseases

For training a model whose goal is to predict the class feature, that is, whether the person has/does not have chronic kidney disease, using the parameters: hemoglobin, red blood cell count, hypertension, and diabetes mellitus, it was found that the best model that minimizes the F1-score result is logistic regression. For which we obtain that the learning curve for F1-score demonstrates that the Logistic Regression model achieves a strong performance, with the training F1-score of 0.9772. The validation F1-score, which is 0.9703, closely aligns with the training score. The standard deviation across cross-validation folds indicates consistent model performance, suggesting a well-fitted model with no significant underfitting or overfitting as the dataset size grows. The accuracy learning curve for Logistic Regression shows the training accuracy initiating near 0.98 and remaining stable across increasing training set sizes, reflecting the model's confidence in the training data. The validation accuracy, which is 0.9625, closely tracks the training accuracy. This convergence, with minimal gaps and tight standard deviation bands, highlights the model's robust generalization capability and absence of overfitting. The log loss learning curve indicates effective learning, with the training log loss decreasing from about 0.45 to 0.1 and stabilizing, signifying a well-optimized model. The validation log loss follows a similar downward trend, dropping to around 0.15, and stabilizing slightly above the training loss. This small gap and the consistent decline suggest that the model generalizes well, with the stabilization indicating it has reached an optimal balance between training fit and validation performance.

Figure 3: A. Learning Curve, F1-score, train, and validation with standard deviation for Logistic Regression.
B. Learning Curve, Accuracy, train, and validation with standard deviation for Logistic Regression.
C. Learning Curve (Log Loss) for Logistic Regression

The degree of the non-linear regression to predict Packed Cell Volume using the features age, hemoglobin, and red blood cell count is 2, using Optuna for this optimization. The Ordinary Least Squares (OLS) regression model was applied to predict Packed Cell Volume, with an R-squared of 0.755 and an adjusted R-squared of 0.748, indicating that approximately 75.5% of the variance is explained by the model. The F-statistic of 105.2 ($p < 0.000$) confirms the model's overall significance. variables $x1$ to $x9$ all statistically significant ($p < 0.05$) except $x9(p = 0.699)$. Diagnostic tests show an Omnibus probability of 0.000, suggesting non-normality, a Durbin-Watson statistic of 1.840 indicating moderate autocorrelation, and a condition number of 8.84, suggesting no severe multicollinearity. The logistic regression model for Diabetes Mellitus shows strong predictive power with a pseudo R-squared of 0.2813. The model converged successfully, with a log-likelihood of -144.67 and a likelihood ratio (LLR) p-value of $2.530e-25$, indicating high significance. both variables, key predictors - age and hemoglobin, are statistically significant. The z-scores and confidence intervals support the reliability of these coefficients. The logistic regression for Hypertension reports a pseudo R-squared of 0.4269, suggesting a good fit with 42.69% of variance explained. The model converged with a log-likelihood of -119.62 and an LLR p-value of $2.138e-38$, indicating strong significance. All three variables, key predictors - age, red blood count, and hemoglobin, are statistically significant, with z-scores ranging from -5.332 to 4.766, and confidence intervals reinforcing their importance.Now, we add new features to the heart disease dataset, which are hemoglobin and red blood cell count as indicated in the methods based on the hypotheses and statistical tests in the first part as well as on the exploration and adjustment that was performed on the NHAMES database along with the addition of normal noise according to the results of the statistical tests. After that, we will now use our trained logistic and nonlinear regression models to add features of packed cell count, Hypertension, and Diabetes mellitus. Finally, we will use the logistic regression model to predict CKD in the heart disease dataset. Now we have all the data we need to test our research hypothesis. A single dataset contains labels whether the patient has/dont have CKD, and the same for heart disease.

A: OLS Regression Results

OLS Regression Results

Dep. Variable:	Packed Cell Volume	R-squared:	0.755
Model:	OLS	Adj. R-squared:	0.748
Method:	Least Squares	F-statistic:	105.2
Date:	Thu, 24 Jul 2025	Prob (F-statistic):	2.29e-88
Time:	16:09:00	Log-Likelihood:	-887.60
No. Observations:	317	AIC:	1795.
Df Residuals:	307	BIC:	1833.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	40.4308	0.376	107.649	0.000	39.692	41.170
x1	-0.7708	0.271	-2.848	0.005	-1.303	-0.238
x2	5.0918	0.338	15.065	0.000	4.427	5.757
x3	1.7668	0.342	5.161	0.000	1.093	2.441
x4	-0.4623	0.182	-2.539	0.012	-0.821	-0.104
x5	0.1188	0.353	0.337	0.737	-0.576	0.813
x6	-0.3340	0.382	-0.874	0.383	-1.086	0.418
x7	-0.2694	0.241	-1.118	0.264	-0.744	0.205
x8	-0.9740	0.560	-1.739	0.083	-2.076	0.128
x9	0.1915	0.364	0.527	0.599	-0.524	0.907

Omnibus:	50.428	Durbin-Watson:	1.840
Prob(Omnibus):	0.000	Jarque-Bera (JB):	110.429
Skew:	0.806	Prob(JB):	1.05e-24
Kurtosis:	5.400	Cond. No.	8.84

B: Logit Regression – Diabetes Mellitus

Logit Regression Results

Dep. Variable:	Diabetes Mellitus	No. Observations:	317
Model:	Logit	Df Residuals:	314
Method:	MLE	Df Model:	2
Date:	Thu, 24 Jul 2025	Pseudo R-squ.:	0.2813
Time:	16:09:00	Log-Likelihood:	-144.67
converged:	True	LL-Null:	-201.31
Covariance Type:	nonrobust	LLR p-value:	2.530e-25

	coef	std err	z	P>|z|	[0.025	0.975]
const	-1.1244	0.172	-6.534	0.000	-1.462	-0.787
x1	-1.1043	0.170	-6.481	0.000	-1.438	-0.770
x2	1.1014	0.197	5.601	0.000	0.716	1.487

C: Logit Regression – Hypertension

Logit Regression Results

Dep. Variable:	Hypertension	No. Observations:	317
Model:	Logit	Df Residuals:	313
Method:	MLE	Df Model:	3
Date:	Thu, 24 Jul 2025	Pseudo R-squ.:	0.4269
Time:	16:09:00	Log-Likelihood:	-119.62
converged:	True	LL-Null:	-208.73
Covariance Type:	nonrobust	LLR p-value:	2.138e-38

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.9732	0.183	-5.332	0.000	-1.331	-0.615
x1	-1.1247	0.241	-4.669	0.000	-1.597	-0.653
x2	1.0086	0.212	4.766	0.000	0.594	1.423
x3	-1.0482	0.288	-3.641	0.000	-1.613	-0.484

Table 1: 1A. Linear regression for Packed Cell Volume CKD feature. 1B. Logistic regression for Diabetes Mellitus CKD feature. 1C. Logistic regression for Hypertension CKD feature.

We now run a Z-test and a non-parametric Binomial test to test the hypotheses and the one-sided result that we hypothesize. The first test is that Most people with heart disease have CKD (proportion > 0.5), and the second test is that Most people with CKD do not have heart disease (the other proportion > 0.5). For the first one, we got for the Z-test a $P - value = 3.5e - 19$, and a corrected P-value that is $7.0e - 19$. And for the Binomial test, a $P - value = 2.1e - 13$, Corrected P-value with the Bonferroni method is $4.2e - 13$. Meaning that indeed, Most people with heart disease have CKD. Now for the second test, we got for the Z-test a $P - value = 0.0009$, and a corrected

P-value that is 0.0017. And for the Binomial test, a $P - value = 0.0013$, Corrected P-value with the Bonferroni method is 0.0025. Meaning that indeed, Most people with CKD do not have heart disease.

We were able to prove that there are indeed dynamics and connections between heart disease and chronic kidney disease. There are dynamics between these two diseases that are hidden from us, but we were able to confirm their existence. This impressive result not only confirms the same research hypotheses for other databases, but also sharpens this connection and shows significance for the fact that a person with heart disease will probably also develop chronic kidney disease as a result, which also stems from the population study we conducted in which we saw that a prominent and significant population about the two diseases is older men. But on the other hand, as we have shown, the connection is not two-way, and if you have chronic kidney disease, this does not necessarily mean that you will develop heart disease. It can also be seen from population studies that chronic kidney disease is usually common in young women/men (under the age of 50), and not like the population with the most prevalent rate for heart disease, older men.

4 Discussions

Our main goal in this article is to prove the existence of connections and dynamics between heart disease and chronic kidney disease. Beyond proving these connections, which we are not the first to claim, we aimed to refine this general statement into a one-sided connection that we hypothesized to exist, namely that heart disease will probably also lead to the development of chronic kidney disease, but not vice versa. This was after we performed comparison tests between populations and first predicted populations that could support this hypothesis. In order to approach this hypothesis, we had to conduct a statistical study of various variables and characteristics of the diseases, as well as use learning models to make predictions and create a uniform database of patients who have documentation of both diseases, to test our hypothesis. During the work, we were indeed able to show the existence of these dynamics between the diseases and answer our research question, which confirmed the macro phenomena that we predicted while observing certain population groups. In this article, we showed the existence of these relationships described above, but we did not conduct a statistical study and analysis using machine learning tools to represent these dynamics using equations, for example, which we would like to continue to study in the future. Another thing, during our work, we made some general assumptions that do not necessarily hold. We assumed that after the statistical study we performed, phenomena that occur for kidney disease may indeed occur with a high probability also in heart patients, for example, low hemoglobin or low red blood cell count. Also, the formula for calculating these parameters, beyond the noise that we added following the statistical tests, is based on previous studies that offer empirical formulas based on different data that purport to present the reality of the normal level for a healthy person. We would also like to study these parameters and infer them from our data. In addition, in our datasets, there was a large amount of missing values, which we solved for and did not replace, for example, a missing value with the mean or median. Therefore, we had a limited data set of 303 samples. We would like to perform these measurements on a larger and more comprehensive database in future research.

References

[1] Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

[2] Student (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25.

[3] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50(302), 157–175.

[4] Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18(1), 50–60.

[5] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83.

[6] Good, P. I. (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Springer Series in Statistics.

[7] Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika*, 55(1), 1–17.

[8] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611.

[9] Anderson, T. W., & Darling, D. A. (1954). A test of goodness-of-fit. *Journal of the American Statistical Association*, 49(268), 765–769.

[10] Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, 83–91.

[11] Fisher, R. A. (1921). On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample. *Metron*, 1, 3–32.

[12] Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621.

[13] Huitema, B. E. (2011). *The Analysis of Covariance and Alternatives: Statistical Methods for Experiments, Quasi-Experiments, and Single-Case Studies*. Wiley.

[14] Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, 69(348), 894–908.

[15] Stevens, J. P. (2002). *Applied Multivariate Statistics for the Social Sciences* (4th ed.). Lawrence Erlbaum Associates.

[16] Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, 39(227), 357–365.

[17] Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg: Friedrich Perthes and I. H. Besser.

[18] Seber, G. A. F., & Wild, C. J. (2003). *Nonlinear Regression*. Wiley-Interscience.

[19] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.

[20] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

[21] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, 30.

[22] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

[23] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.

[24] Chew, N. W. S., Tay, W. T., Richards, M., et al. (2023). Association Between Diabetes, Chronic Kidney Disease, and Outcomes in People With Heart Failure From Asia. *The Lancet Regional Health – Western Pacific*, 37, 100831. `https://doi.org/10.1016/j.lanwpc.2023.100831`

[25] Salive, M. E., Cornoni-Huntley, J., Guralnik, J. M., et al. (1992). Anemia and hemoglobin levels in older persons: Relationship with age, gender, and health status. *Journal of the American Geriatrics Society*, 40(5), 489–496.

[26] Beutler, E., & Waalen, J. (2006). The definition of anemia: What is the lower limit of normal of the blood hemoglobin concentration? *Blood*, 107(5), 1747–1750.

[27] Guralnik, J. M., Eisenstaedt, R. S., Ferrucci, L., Klein, H. G., & Woodman, R. C. (2004). Prevalence of anemia in persons 65 years and older in the United States: Evidence for a high rate of unexplained anemia. *Blood*, 104(8), 2263–2268.

[28] Rappoport, N., et al. (2015). Red blood cell parameters: Association with age and sex. *PLoS One*, 10(7), e0132788. `https://pubmed.ncbi.nlm.nih.gov/26119080`

[29] Hall, J. E., & Guyton, A. C. (2020). *Guyton and Hall Textbook of Medical Physiology* (14th ed.). Elsevier.

[30] Walker, H. K., Hall, W. D., & Hurst, J. W. (Eds.). (1990). *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Boston: Butterworths.