

# **Impact of Dietary Recommendations Shift on Chronic Kidney Disease**

Sayed Ahmed, Dr. Youcef Derbal

## **Abstract**

Chronic Kidney Disease (CKD) leading to End-Stage Renal Disease (ESRD) is very prevalent today. Over 37 millions of Americans have CKD. CKD/ESRD and interrelated diseases cause a majority of the early deaths. Many research studies have investigated the effects of drugs on CKD. However, less attention has been given to the study of the dietary patterns on CKD. This research study has uncovered significant correlations between dietary patterns and CKD mortality as well as identified diagnostic markers for CKD such as the Albumin to Creatinine Ratio (ACR). In this project, Dietary surveys from NHANES, and CKD Mortality dataset from USRDS were utilized to study the correlation between dietary patterns and morbidity of CKD patients. Principal Component Analysis and Regression were utilized to find the effect. Machine Learning Approaches including Regression, and Bayesian were applied to predict ACR values. Grains, Other Vegetables showed positive correlations with Mortality whereas Alcohol, Sugar, and Nuts showed negative correlations. ACR values were not found strongly correlated with dietary patterns. For ACR value prediction, 10 Fold Cross Validations with Polynomial Regression showed 95% accuracy.

### **Keywords:**

CKD, ESRD, Dietary Patterns, Mortality, ACR, Polynomial Regression, Bayesian

## **1 Introduction**

Chronic kidney disease (CKD) is very prevalent in today's world and CKD incidents are continually

increasing whereby 10 to 13% of the US population are affected by Chronic Kidney Disease. CKD/ESRD and other interrelated diseases such as Hypertension, Heart Diseases, and Diabetes cause the majority of early deaths [31]. In addition to kidney failure, CKD is also a major cause of death from stroke, and heart diseases. On the other hand, hypertension and diabetes are also major causes of CKD. CKD is not reversible, progressive and gradually reduces kidney function.

Primary causes of CKD include High Blood Pressure, and Diabetes. Other causes include infection, kidney stones, genetics, genetical polycystic kidney disease, certain foods and food habits, pain killers, and drug usage/abuse. As CKDs are neither curable nor reversible controlling diabetes and blood pressure with or without medication can slow the progress of CKDs in most cases. As Kidneys filter waste products and our diet produces those waste products controlling diet have an effect on how much work kidney has to perform, and how well kidneys will function. Studies show that drugs as well as lifestyle choices (food, diet, exercise) can prevent CKD, slow the progression of CKD [29], delay dialysis and kidney transplantation; consequently can prevent early deaths. Although there are many studies on the effect of drugs to control CKD and related complications, there are few studies on the effect of diets, dietary patterns, and lifestyles [29]. There are studies on how controlling nutrients/chemicals in food items can help to prevent or to slow the progression of CKD. CKD patients are also provided recommendations on certain chemicals or food items. However, adhering to the recommended amount of nutrients and/or food item is challenging. Hence, there is an emerging trend where the effect is studied utilizing dietary patterns with food groups and food

subgroups rather than nutrients/chemicals in food or individual food item. This research analyzes the effect of dietary patterns, using food groups and food subgroups, on the mortality and survival of CKD patients. Also studied in this research is the correlation between dietary patterns and ACR. Additionally, ACR values are predicted based on a dataset utilizing machine learning approaches.

## 1.1 How CKD is Identified and Measured

CKD is identified with one of two measures such as a blood test named Glomerular Filtration Rate (GFR) or a urine test named Albumin to Creatinine Ratio (ACR). ACR values less than 30 indicates no CKD or mild CKD. ACR values between 30 and 300 indicate moderate CKD. ACR values  $>300$  indicates severe CKD. Patients are diagnosed with a CKD disease if the ACR values persist within the above ranges for three months. GFR is measured in  $\text{ml/min}/1.73 \text{ m}^2$ . CKDs as measured with GFRs are described in stages such as Stage 1 with normal or high GFR (GFR greater than or equal to  $90 \text{ mL/min}$ ), Stage 2 with Mild CKD (GFR =  $60\text{-}89 \text{ mL/min}$ ), Stage 3A with Moderate CKD (GFR =  $45\text{-}59 \text{ mL/min}$ ), Stage 3B with Moderate CKD (GFR =  $30\text{-}44 \text{ mL/min}$ ), Stage 4 with Severe CKD (GFR =  $15\text{-}29 \text{ mL/min}$ ), Stage 5 with End Stage CKD (GFR  $<15 \text{ mL/min}$ ) [5]. At stage 5, patients lose complete kidney function. At this point, patient survival will require either dialysis or organ transplantation.

## 2 Literature Review

Kidney patients are commonly given dietary advice about intake of individual nutrients, chemicals, food items, or about their whole eating patterns.

However, dietary recommendations are challenging to adhere to for the majority of the patients [2]. Also, there is limited evidence that adherence to such advice prevents clinical complications [23]. Hence, studying the whole dietary patterns rather than single nutrient or food group restrictions is an emerging trend for CKD/ESRD patient diets [2] [24-26]. This is also easier to adhere to. There are several studies on analyzing the relation between dietary patterns and clinical outcomes for CKD patients [3, 4, 5, 6, 7, 8, 9, 26].

Chen et al [3] studied the association between plant protein intake and mortality in CKD. In the study higher plant protein ratio was found to cause lower mortality for CKD patients in stage 3 or higher (  $\text{eGFR} < 60 \text{ ml/min}/1.73 \text{ m}^2$  ) though not for others (stage 1 and 2) [3]. This study primarily used statistical methods and Regression Models such as Cox regression models to find the association [3]. Hao-Wen et al [26] studied the association between vegetarian diets and CKD. The study found that vegetarian diets including vegan and ovo-lacto vegetarian diets were possible protective factors. The study utilized The multivariable logistic regression analysis [26].

Gutiérrez et al [4] studied 5 empirically derived dietary patterns such as "convenience" (Chinese and Mexican foods, pizza, and other mixed dishes), "plant-based" (fruits and vegetables), "sweets/fats" (sugary foods), "Southern" (fried foods, organ meats, and sweetened beverages), and "alcohol/salads" (alcohol, green-leafy vegetables, and salad dressing) [4]. The study found that dietary pattern rich in processed and fried foods was associated with higher mortality in persons with CKD. On the other hand, a diet rich in fruits and vegetables was found to be protective [4].

Huang et al. [5] studied whether Mediterranean

diet can preserve kidney function along with maintaining favorable cardiometabolic profile with reduced mortality risk for individuals with CKD. The study found that adhering to Mediterranean diet has a lower likelihood of having CKD in elderly men. The study also found that a greater adherence to this diet can improve survival for CKD patients [5]. Huang et al [5] in the above study, used unpaired t test, nonparametric Mann–Whitney test, or chi-square test as appropriate for Comparisons between CKD and non-CKD men. To evaluate the association of Mediterranean diet with the presence of CKD, Crude and multiple adjusted logistic regression models were fitted. All tests were two-tailed, and  $P < 0.05$  was considered significant [5]. One aspect of Muntner et al [6] study was to find out how Life’s Simple 7 factors (Smoke, Activity, BMI, Diet, Blood Pressure, Cholesterol, and Glucose) affect in getting ESRD. For each factor, participants were assigned to one of three levels of compliance to the recommendations. The three levels are ideal, intermediate, and poor. The study shows that people who have ideal scores in more of these factors have lower likelihood of getting ESRD. This study utilized Cox proportional hazards models. Adjustment were made for age, race, sex, stroke-based geographic region of residence, income, education, and history of stroke or coronary heart disease [6].

Ricardo et al [7] studied the association of death to healthy lifestyles especially in relation to CKD. The study found that adherence to healthy lifestyles was associated with lower risk for mortality in CKD patients. In this study, to determine the association between a healthy lifestyle and survival among individuals with CKD, Cox proportional hazards models were used while also adjusting for important covariates. Stratified survival analyses

by eGFR and UACR was performed for Sensitivity analyses [7]. Suruya et al. [8] studied dietary patterns in hemodialysis patients in Japan and researched associations between dietary patterns and clinical outcomes. The study found that patients with unbalanced diet were more likely to have adverse clinical outcomes. Hence, such patients when in addition to portion control, maintains a well-balanced diet especially for the food groups meat, fish, and vegetables will have less adverse clinical outcomes [8]. Suruya et al [8] utilized a principal components analysis (PCA) with Promax rotation to reduce to a smaller set of food groups for analysis. PCA was used to find food groups eaten with equal frequencies [8]. Cox regression model was used for the analysis with multiple models where each model had a different combination of covariants [8].

Another study by Ricardo et al [9] estimated the degree of adherence to a healthy lifestyle that decreases the risk of renal and cardiovascular events among adults with chronic kidney disease (CKD). The study found that adherence to a healthy lifestyle was associated with lower all-cause mortality risk in CKD. The greatest reduction in all-cause mortality was related to nonsmoking [9]. This study by Ricardo et al [9], to compare categorical and continuous variables used Chi-squared and analysis of variance tests respectively. To examine the association between healthy lifestyle and outcomes, Cox proportional hazards models were used. Death was treated as a censoring event. Three nested Cox proportional hazards models were fitted and were adjusted sequentially for potential explanatory variables [9]. G. Asghari et al [27] studied the association of population-based dietary pattern with the risk of incident CKD. The study concluded that high fat and high sugar diet pattern is associated with

significantly increased (46%) odds of incident CKD where a lacto-vegetarian diet can be protective of CKD by 43%. The study utilized multivariable logistic regression to calculate odds ratio for the association.

Overall, most of the reviewed studies primarily used direct clinical data of patients for several years and applied statistical analysis primarily. This research primarily utilized public datasets from CDC, USRDS. One of the studies above utilized the dietary pattern data from CDC and NHANES like this study. However, this study will differ in the methodology, exploration, and analysis. This study is finding relations between datasets from multiple sources and is focused on finding patterns and relations in general population than specific/selected individuals. Most of the studies above utilized primarily statistical methods and sensitivity analysis where primarily regression models especially Cox regression models were used. In a couple of cases, Principal Component Analysis (PCA) was used. This research primarily utilized public datasets from CDC, USRDS, Health.gov. For association and prediction, PCA, Regression, and several machine learning approaches are utilized. For food groups and subgroups, we utilized the USDA categorization. Recommended amounts for food groups provided by CDC/Health.gov is used. Additionally, ACR values are predicted based on datasets by CDC such as NHANES dietary intake survey, and laboratory data on ACR. For the prediction, machine learning approaches are used. The machine learning approaches used for ACR value prediction are: Regression and Bayesian.

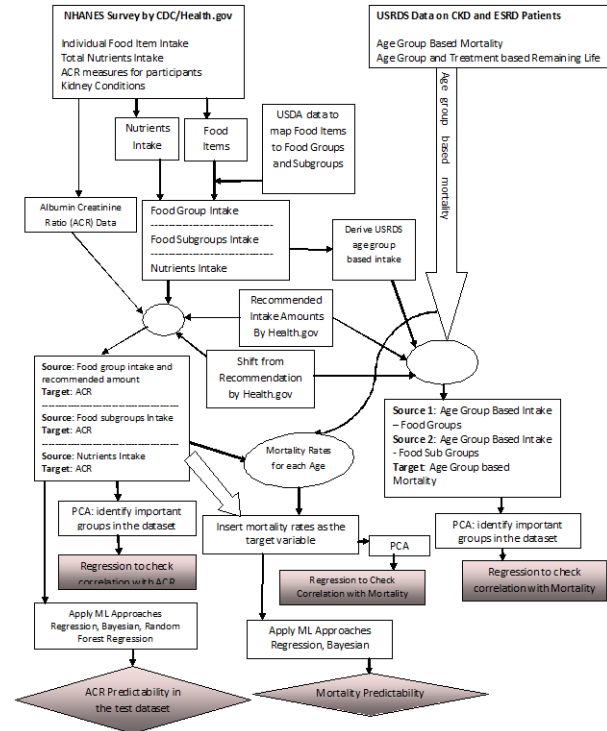


Figure 1: Methodology in a Diagram

### 3 Methodology

The research methodology is summarized by the diagram given in Figure 1. The primary purpose of this research is to assess the effect of dietary patterns on the ACR levels, survival and morbidity for chronic kidney disease (CKD)/End Stage Renal Disease (ESRD) patients.

#### 3.1 Methodology Overview

A dataset released by CDC/Health.gov with the dietary habits and ACR values for 10,000 individuals are studied. Also, age group based mortality and survival of CKD/ESRD patients provided by USRDS are studied. Afterwards, utilized Principal Component Analysis to identify the most important food groups and subgroups affecting the ACR value and the Mortality/Survival. Statistical regression and factor analysis was then applied to understand the correlation between ACR, and Mortality/Survival

to dietary patterns. Machine learning approaches including Regression, Polynomial Regression, and Bayesian with or without 10 fold cross validations are applied on the datasets to understand if dietary patterns can be used to predict ACR values and mortality.

In addition to studying the association of CKD mortality and ACR values with ratios from the recommended high, we have explored the potential impact of diet recommendation shift [11], on the incidence of CKD in the general population.

### 3.2 Study Selection

For dietary patterns, CKD measures (such as Albumin to Creatinine Ratio - ACR), and Kidney condition measures, a dataset from the National Health and Nutrition Examination Survey on dietary habits conducted by the CDC [10] was used. The survey has data from 1996 to 2016 [10]. This study primarily utilized data for 2015-2016. The survey recorded 24 hours individual food item intake amount. Two surveys were taken within 3 to 10 days apart. Each survey provided food item intake amounts in a day and recorded the diet style as well as diet-restrictions. Individual food items are represented using USDA food code. The survey also provided total nutrients data. CDC also released examination, laboratory, demographics, and other related data for those participants. This study explored and utilized examination data such as Kidney Condition data, laboratory data such as ACR data & Blood Pressure data, and demographics data for age.

For mortality and survival information, dataset from the United States Renal Data System (USRDS) on CKD and ESRD [16, 17] were utilized. “USRDS investigates the transition of care from CKD to ESRD and end-of-life care for those

with advanced kidney disease” [19]. USRDS also releases data on the Incidence, Prevalence, Patient Characteristics, and Treatment Modalities on CKD, and ESRD patients. USRDS reports survival and mortality using metrics such as Mortality rates, Total Mortality Count, 90 day survival for dialysis and/or transplantation patients, 10 year survival for dialysis and/or transplantation patients, Average Expected remaining lifetime with or without pre-condition and treatment options used. The data are either aggregated or patient specific detail data. However, only aggregated data are public where patient specific data access requires special request and permission. This research utilized only the public dataset i.e. age-group based aggregated data. In couple of experiments, aggregated age-group based data from USRDS was mapped to specific age for NHANES data for each age and participants.

The dietary survey data (NHANES) represented the food items taken by the participants using USDA food codes [14, 12, 13]. Hence, USDA food codes [14, 12, 13] are used to assign food groups and subgroups to the NHANES [10] survey data to properly group/subgroup the dietary intake of the participants with some customizations.

### 3.3 Data Synthesis

NHANES survey data as provided for two days are averaged to get the intake amount for one day. Both individual food item data and nutrients intake data are averaged. USDA food codes are used to map food items to food groups and subgroups.

ACR and Kidney condition data for each individual are merged with the averaged food groups, subgroups, and nutrients data. This data was further complemented with the food group recommendations data from health.gov, and

mortality rate by age from the USRDS. ACR values and mortality rate are used as the target variables. PCA was applied to find out important food groups and subgroups while regression was applied to find potential associations with ACR and Mortality. Linear Regression, Polynomial Regression, Random Forest Regression, Bayesian prediction with or without 10 fold cross validations were applied to study the predictability of ACR Values and Mortality in the test dataset.

In another mortality experiment, the above synthesized datasets were aggregated for USRDS age groups to calculate average food group/subgroup intake by age groups.

## 4 Results

Associations of Food Groups, Food Subgroups, Food Nutrients with CKD mortality as discovered by the experiments above are provided in this section. Outcome of ACR value prediction in the test data are also provided.

### 4.1 Food Groups, Food Subgroups, Food Nutrients and CKD Mortality

#### Food Groups and Mortality

Experiments (Set 1 ) with aggregated NHANES and USRDS data to find associations between food groups and CKD mortality using PCA and Regression show that Grains (-0.84) and Fruits (-0.43) have negative correlations with CKD mortality i.e. mortality is high for the patients who took significantly lower amount of Grains and Fruits than recommended amounts. Data exploration (plots below) also reflects the negative relation. As the correlation for fruits is -0.43 i.e. not very high, hence, Fruits can be thought of

mildly/moderately associated.

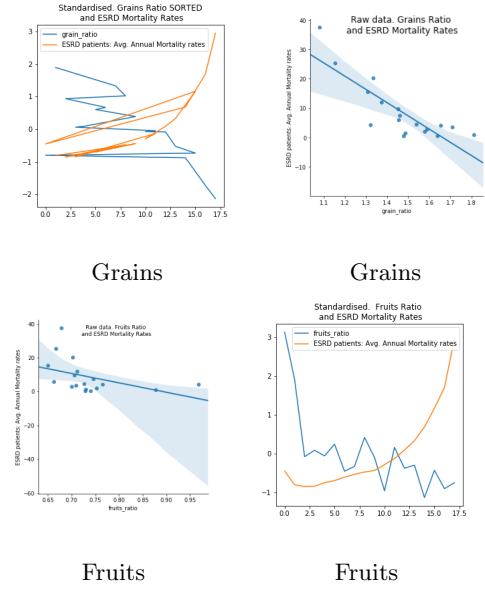


Figure 2: Food Groups and Mortality - Negative Correlations

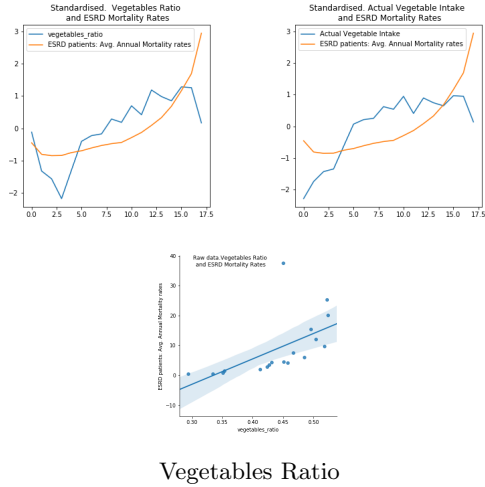


Figure 3: Food Groups and Mortality - Positive Correlations (Vegetables)

Vegetables show positive (0.58) correlation i.e. mortality is high for the patients who took more vegetables. The correlation of 0.58 does not provide a very strong conclusion. Data shows such correlations in older adults. Even though ratios of food intake amount to high end of recommended amounts (actual intake amounts also show positive relation) were used; age might

have biased the correlation. This does not show conformity with the general recommendation to take more vegetables for CKD patients. However, as experiments with food subgroups show that vegetable subgroups such as Other vegetables (0.68), Red and Orange vegetables (0.55), and Starchy vegetables (0.44) show positive correlations that have an impact on the Vegetable group correlation. Although more vegetable intake is a general recommendation for CKD patients, certain vegetables with more carbohydrates (and sugars) such as Starchy Vegetables as well as Vegetables with more Potassium and Calcium such as Tomatoes, and Spinach are recommended to be taken at a lower amount. For diabetes induced CKD, starchy vegetables are not highly recommended in general. Considering the subgroups, the moderate positive correlation as this study found for vegetables food group is in conformity with the general recommendations for CKD patients.

Experiments (Set 2) with aggregated NHANES and USRDS data to find associations between food subgroups and CKD mortality using PCA and Regression show that Other vegetables (0.68), Red and orange vegetables (0.55), and Starchy vegetables (0.44) have positive correlations with mortality i.e. mortality is low when the intake amounts are low, and mortality is high when intake amounts are high. Data Exploration plots as given below also show these positive relations.

Food subgroups such as Alcoholic Beverages (-0.79), Added Sugars/Sugars and Sweets (-0.64), Whole Grains (-0.61), and 'Nuts, Seeds, and Soy Products' (-0.55) show the most negative correlations with CKD mortality. Data exploration also shows negative correlations as shown in the charts below. These outcomes are also consistent with current knowledge except for

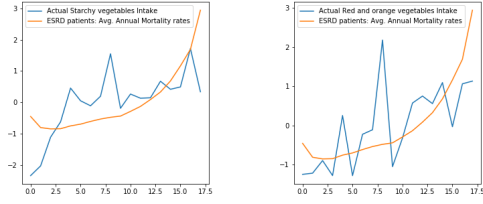
Sugars. Prevalence of Stage 3 CKD is lower in Alcohol Drinkers than non-drinkers [2, 49], Nuts being Phosphorus rich and Whole Grains being Potassium rich are detrimental to CKD patients and can cause higher mortality when taken in higher quantities.

### **Mortality Study with Non-aggregated Data**

For experiments (Set 7) where mortality rates based on ages were used for each NHANES survey row (i.e. not aggregated), the following food subgroups show more positive correlations than others: Fats, Eggs, Other vegetables, Nonalcoholic beverages, White potatoes and Puerto Rican starchy vegetables, Tomatoes and tomato mixtures, Oils, Deep-yellow vegetables respectively. In the study, the following subgroups showed more negative correlations than others: Grain mixtures, Frozen plate meals, Soups, Crackers and Salty snacks from grain products, Milks and milk drinks, Sandwiches with Meat, Poultry, and fish, Poultry. Although the correlation numbers in this study are very low, the positive and negative correlations are consistent with current knowledge [48, 49, 50, 51, 52, 53, 54, 55, 56, 57].

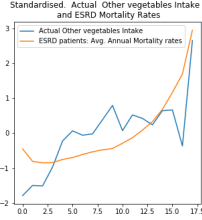
### **Food Groups, Food Nutrients, and Albumin to Creatinine Ratio (ACR) Association**

The experiments (Set 3, 4, and 5) using PCA and Regression showed negligible correlation between ACR and food groups and subgroups intake. However, some food groups and/or nutrients such as Dairy, and 'Sugars, Sweets, and Beverages' have higher and positive though negligible (0.02) effect than the others where Fruits (-0.01) showed negative effect. For nutrients, Polyunsaturated fatty acids (-0.02), and Iron (-0.02) have negative correlations where Choline (0.02) showed better positive correlation than others. Findings for Choline matches with medical knowledge [48]. As the correlations are not significant further analysis



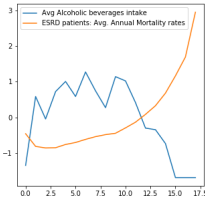
Starchy Vegetables

Red and Orange Vegetables

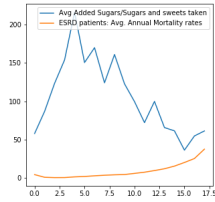


Other Vegetables

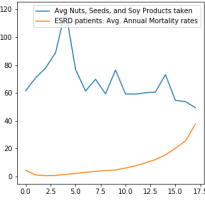
Figure 4: Food Subgroups and Mortality - Positive Correlations



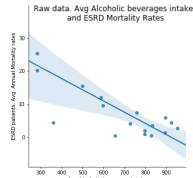
Alcohol Intake



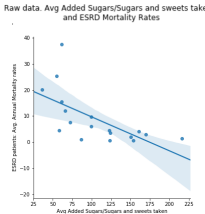
Sugar



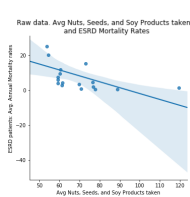
Nuts, Seeds



Alcohol



Added Sugar



Nuts, Seeds

Figure 5: Food Subgroups and Mortality - Negative Correlations

can be done on the data especially for food groups and nutrients that are found important (using PCA) in the data as provided below:

Table 1: Outcome when ACR Values are used as the Target

Data Normal ized	Target	Approach	MSE Train	MSE Test	Accuracy
No	ACR Value	10 Fold Cross Validation Polynomial Regression			-0.957 cross val score
No	ACR Value	Polynomial Bayesian with Cross Validation			-0.682 cross val score
No	ACR Value	Polynomial Regression	90965	52946	-0.579 r2 score on test data
No	ACR Value	Bayesian on Polynomial fit	93047	47431	-0.414 r2 score on test data

Dairy, Fats, oils, and salad dressings', Fruits, Grains, Protein, Sugars, sweets, and beverages', Vegetables, Avg energy kcal, avg protein gm, avg carbohydrate gm, avg total fat gm, avg total saturated fatty acids gm, avg total monounsaturated fatty acids gm, avg total polyunsaturated fatty acids gm, avg lutein zeaxanthin mcg, avg thiamin vitamin B1 mg, avg riboflavin Vitamin B2 mg, avg Niacin mg, avg Calcium mg, avg Phosphorus mg, avg Magnesium mg, avg Iron mg, avg Zinc mg, avg Copper mg, avg Sodium mg, avg Potassium mg, avg Selenium mcg, Hexadecenoic gm, Octadecenoic gm

## 4.2 Food Subgroups and Albumin to Creatinine Ratio (ACR) Association

The experiments showed 'Milk desserts, Sauces, Gravies' (0.22), and Alcoholic Beverages (0.087)



have more positive correlations with ACR than the other food subgroups i.e. taking more of these food subgroups results higher ACR values. Research by Uehara et al. [55] also shows that excessive Alcohol consumption can cause Proteinuria/Albuminuria (high ACR). Nettleton et al. [56] found that high fat dairy can be linked to high ACR values where low-fat dairy is not strongly linked to high ACR values. However, the correlation as this research found is very low. Low values might still explain a correlation where ACR values might depend on other factors in together than only these food subgroups. Fruits and juicy baby foods show negative correlation (-0.04) though not significant i.e. taking high amount does not increase ACR values that matched with current knowledge [57].

### 4.3 Predictability of ACR values based on Dietary patterns

Experiments (Set 6 ) were conducted to predict ACR values from the dietary intake patterns data. Machine Learning (ML) Approaches of Regression, Polynomial Regression, Random Forest Regression, Bayesian prediction with or without 10 fold cross validations were applied on Food Subgroups intake dataset. Only the food subgroups that were found to be important using PCA were used for the ML approaches.

#### Target Variables

Absolute ACR values and ACR Categories were used as the target variables. For ACR category, ACR <30 is assigned to class 0 (i.e. no ckd), and ACR >30 is assigned to class 1 (CKD). ACR values less than 30 indicate no CKD, where ACR values between 30 and 300 indicate moderate CKD. ACR >300 is considered severe CKD.

#### Outcome when ACR Values are used as the Target

Table 2: **Outcome when ACR Category is used as the Target**

<b>Data Normal</b>	<b>Target</b>	<b>Approach</b>	<b>Train Confusion Matrix</b>	<b>Test Confusion Matrix</b>
No	Category	Linear Regression	6927, 87% (6032,0, 895,0)	770, 88% (692, 0, 78, 0)
Yes	Category	Linear Regression	6927, 87% (6032,0, 895,0)	770, 88% (692, 0, 78, 0)

The best test set accuracies were found using approaches such as: 10 Fold Cross Validations with Polynomial Regression (95%), Polynomial Bayesian with 10 fold Cross Validations (68%), Polynomial Regression (57%), Bayesian on Polynomial fit (41%), Cross Validation with Polynomial Random Forest Regression (21%). A list of the best performing approaches and the outcome are provided in Table 1 below.

#### Outcome when ACR Category is Used as the Target

After regression,  $y > 0.5$  is assigned to category 1 (CKD), others were assigned to category 0. Test accuracies are 88%. Format for confusion matrix used in the Table 2 below is: (Total, % Correct : [ TP, FN, FP, TN] ). The high prediction accuracies might relate to the fact that only 10 to 13% of the population has ACR >30. However, as cross validations also show high accuracies, it can be concluded that ACR values can be well predicted using Machine Learning approaches especially with 10 Fold Cross Validation Polynomial Regression having accuracy 95%.

## 5 Conclusions

Chronic Kidney Disease (CKD) leading to End Stage Renal Disease (ESRD) is very prevalent today; treatment facilities for dialysis, and donors for organ transplantation are limited. Consequently, many patients die waiting for proper treatment (recent news on USA, Bloomberg). Majority of the studies focused on drugs to control CKD progression where some studies focused on diets, Nutrients, and Food Items. Controlling CKD using changes to dietary patterns can be beneficial to both the patients and the economy. Hence, this study focused on the effect of dietary patterns on CKD mortality as well as on a CKD measure named Albumin to Creatinine Ratio (ACR). PCA and Regression are used to study the association between ACR and CKD mortality/survival. Additionally, regression models are trained to predict ACR values from dietary patterns. Grains, Other Vegetables showed positive correlations with Mortality whereas Alcohol, Sugar, and Nuts showed negative correlations. ACR values were not found strongly correlated with dietary patterns. Overall, the results of this research matched the findings of other studies and current knowledge with few exceptions. Machine Learning approaches showed that ACR values could be predicted in the test dataset with high accuracy; the best performing approach was 10 Fold Cross Validation for Polynomial Regression (95%).