|  |  |  |
| --- | --- | --- |
| **Title** | | The Effect of Potassium on Chronic Kidney Disease |
| **Investigator(s)** | | Tina Davenport, PhD |
| **Biostatistician(s)** | | Brandon Sucher |
| **Original Creation Date** | | September 1, 2021 |
| **Submission Deadline(s)** | | Monday, November 22, 2021 at 11:59pm |
| **Activity Log** | 08/30/2021 File created | |

# Study Overview

Chronic kidney disease (CKD) affects approximately 37 million adults in the United States, with millions more at increased risk of developing CKD. CKD is often accompanied by other comorbid conditions, particularly hypertension, cardiovascular disease, and diabetes. In addition to addressing these known contributors to CKD, identifying and intervening on other modifiable risk factors could help slow the increase in CKD prevalence. Potassium (K) has been shown to be associated with lower risk of diabetes and cardiovascular outcomes, but its potential impact on CKD is unclear. Therefore, the goal of this study is to investigate the relationships between K and CKD among a cohort of participants in the National Health and Nutrition Examination Survey (NHANES) program, a set of studies designed to assess the health and nutritional status of adults (and children) in the United States.

## Study Aims

### Aim 1

Investigate the association between serum K and CKD.

### Aim 2

Investigate the association between dietary K and CKD.

### Aim 3

Investigate the association between serum K and kidney function.

### Aim 4

Investigate the association between dietary K and kidney function.

## Research Hypotheses

### Hypothesis 1

Higher serum K will be associated with negative CKD status.

### Hypothesis 2

Higher dietary K will be associated with negative CKD status.

### Hypothesis 3

Higher serum K will be associated with better kidney function.

### Hypothesis 4

Higher dietary K will be associated with better kidney function.

# Study Population

The dataset comes from the National Health and Nutrition Examination Survey (NHANES) program, which has been [described in detail](https://www.cdc.gov/nchs/nhanes/about_nhanes.htm) previously. A subset of 350 participants from the 2017-2018 survey were used as the initial sample for this study.

## Inclusion/Exclusion Criteria

Only those with complete data on CKD status and serum K will be included in the analyses.

For each aim separately, those with additional missing data will be excluded as needed to perform the analyses.

## Data Acquisition

|  |  |
| --- | --- |
| Study design | Retrospective, cohort, cross-sectional, observational study |
| Data source/how the data were collected | NHANES 2017-2018; data collection is well documented on the [NHANES website](https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017) |
| Contact information for team member responsible for data collection/acquisition | [Tina.Davenport@duke.edu](mailto:Tina.Davenport@duke.edu) for data acquisition |
| Date or version (if downloaded, provide date) | 2017-2018 NHANES data downloaded on April 21-22, 2021. |
| Data transfer method and date | Data downloaded directly from the [NHANES website](https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017) on April 21-22, 2021.  Analysis dataset finalized on August 27, 2021 |
| Where dataset is stored |  |

# Outcomes, Exposures, and Variables of Interest

## Primary Outcome(s)

The primary outcome is CKD. We will treat CKD as a binary variable, and we will also use eGFR as a proxy for general kidney function.

|  |  |  |
| --- | --- | --- |
| Outcome | Description | Specifications / Notes |
| CKD | Chronic kidney disease  No CKD is defined as eGFR≥60 AND ACR <30.  CKD is defined as eGFR <60 OR ACR ≥30. | A derived binary variable where  0= no CKD  1= CKD |
| eGFR | Estimated glomerular filtration rate (mL/min/1.73m2)  This was calculated using CKD-EPI with the race coefficient omitted. <https://www.kidney.org/content/ckd-epi-creatinine-equation-2009>  A valid range is 0≤ eGFR < 150. | A continuous variable with values ranging from 8.62 to 142.03. |

## Secondary Outcomes

None.

## Primary Exposure(s)

|  |  |  |
| --- | --- | --- |
| Exposure | Description | Specifications |
| serumK  serumK2cat | Serum potassium (mmol/L)  A valid range is 3 ≤ SerumK < 6  Serum K can be dichotomized into low-normal (3 ≤ K ≤ 4) and normal (K > 4) | serumK is a continuous variable with values ranging from 3 to 5.4  serumK2cat is a dichotomous variable where  0 = Low-normal  1 = Normal |
| dietK1000  dietK2cat | Dietary potassium (mg/1000kcal)  Computed as the average potassium intake from food over two days, and scaled to a 1000 calorie diet  Diet K can be categorized into inadequate intake (K < 1534), borderline adequate intake (1534≤ K < 2238), and adequate intake (K ≥2238) | dietK1000 is a continuous variable with values ranging from 314.5 to 3076.4  dietK2cat is an ordinal variable where  1 = Inadequate intake  2 = Borderline adequate intake  3 = Adequate intake |

## Other Variables of interest

|  |  |  |
| --- | --- | --- |
| Variable | Description | Specifications |
| SEQN | Participant ID number | A discrete variable between 93716 and 102914 |
| Age | Age (years)  Ascertained at the time of screening. Individuals 80 and over were topcoded at 80 years.  A valid range is 18 ≤ Age ≤ 80 | A discrete variable with a range of 18 to 80. |
| Sex | Sex of the participant | A dichotomous/binary variable where  F= female  M= Male |
| Race | Reported race and Hispanic origin information | A nominal variable where  1= Mexican American  2= Other Hispanic  3 = Non-Hispanic White  4 = Non-Hispanic Black  6 = Non-Hispanic Asian  7 = Other / Multi-Racial |
| Marital  Married 2cat | Marital status | Marital is a categorical variable where  1= Married  2= Widowed  3= Divorced  4= Separated  5= Never married  6= Living with partner  Married2cat is a dichotomous variable where  0= Not married or living with partner (2, 3, 4, 5)  1= Married or living with partner (1, 6) |
| Educ  Educ2cat | Education  Ascertained by the question “What is the highest grade or level of school you have completed or the highest degree you have received?” | Educ is an ordinal variable where  1 = Less than 9th grade  2 = 9-12th grade (no diploma)  3 = High school graduate/GED or equivalent  4 = Some college or AA degree  5 = College graduate or above  Educ2cat is a dichotomous variable where  0 = HS or less (1, 2, 3)  1 = Some college or more (4, 5) |
| Income  Income5cat | Total household income  Reported as a range value in dollars | Income is an ordinal variable where  01 = $0 to $4,999  02 = $5,000 to $9,999  03 = $10,000 to 14,999  04 = $15,000 to $19,999  05 = $20,000 to $24,999  06 = $25,000 to $34,999  07 = $35,000 to $44,999  08 = $45,000 to $54,999  09 = $55,000 to $64,999  10 = $65,000 to $74,999  12 = $20,000 and over  13 = Under $20,000  14 = $75,000 to $99,999  15 = $100,000 and over  77 = Refused  99 = Don’t know  Income5cat is an ordinal variable where  0: Under 20k  1: 20k to <45k  2: 45k to <75k  3: 75k and over  98: Refused/DK |
| BMI | Body mass index (kg/m2)  A valid range is 15 ≤ BMI < 70 | A continuous variable with values ranging from 16.6 to 84.4. |
| Smoking | Smoking status  Derived from two questions: 1) “Have you smoked at least 100 cigarettes in your entire life?” and 2) “Do you now smoke cigarettes?”.  Current smoker is defined as yes/yes. Former smoker is defined as yes/no. Never smoker is defined as no/no. | An ordinal variable where  0 = Never  1 = Former  2 = Current |
| Alcohol12m | Alcohol use  Ascertained by the question “During the past 12 months, about how often did you drink any type of alcoholic beverage?” | An ordinal variable where  0 = None  1 = A few times  2 = Monthly  3 = Weekly |
| SBP | Systolic blood pressure (mm Hg)  Computed as the average of 4 readings.  A valid range is 70 ≤ SBP < 180 | A continuous variable with values ranging from 78.67 to 203.33 |
| DBP | Diastolic blood pressure  Computed as the average of 4 readings.  A valid range is 40 ≤ DBP < 120 | A continuous variable with values ranging from 0 to 104.67 |
| ACR | Urinary albumin to creatinine ratio (mg/g)  A valid range is 0 ≤ ACR < 6000 | A continuous variable with values ranging from 0.54 to 11055.12 |
| A1C | Glycohemoglobin (%)  A valid range is 4 ≤ A1C < 11 | A continuous variable with values ranging from 4.7 to 16.2 |
| DM | Diabetes status  Glucose control is defined as  no self-report of DM/preDM AND (FPR < 100 or non-FPG < 200) AND A1C < 5.7 AND no DM medication  Prediabetes is defined as  self-report prediabetes OR 100 ≤ FPG < 126 OR 5.7 ≤ A1C < 6.5  Diabetes is defined as  self-reported diabetes OR self-reported borderline diabetes OR FPG ≥ 126 mg/dL OR non-FPG > 200 OR A1C ≥ 6.5% or diabetes medication | An ordinal variable where  0 = GC/IND  1 = preDM  2 = DM |
| Sleep | Sleep (hours)  Number of hours usually sleep on weekdays or workdays | A continuous variable with values ranging from 2 to 14 |
| PHQ9  PHQ5cat | Depression score  Derived from a 9-item instrument called the Patient Health Questionnaire (PHQ) to determine the frequency of depression symptoms over the past 2 weeks.  A valid range is 0 ≤ PHQ ≤27, where a higher score means more depression. PHQ score can be categorized into minimal (0-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27). | PHQ9 is a discrete variable with values ranging from 0 to 25.  PHQ5cat is an ordinal variable where  0 = None-minimal (0-4)  1 = Mild (5-9)  2 = Moderate (10-14)  3 = Mod Severe (15-18)  4 = Severe (21-25) |

# Statistical Analysis Plan

### Exploratory Analyses and CONSORT Numbers

We will describe any oddities or unexpected aspects of the data. We will apply our inclusion/exclusion criteria outlined above and describe how our analytic sample was obtained.

We will then investigate the distributions of continuous variables via histograms, density plots, and boxplots, and categorical variables via bar plots and crosstabs as needed. We will explore the relationships between the outcomes and exposures via crosstabs, histograms, boxplots, and scatterplots (with LOESS) as appropriate. We will check assumptions, identify outliers, and explore missingness as needed.

### Graphical Displays

We will present relevant plots from exploratory analyses.

### Table 1

Participants’ demographic and clinical characteristics will be presented across all subjects and stratified by serum K value. We will do this by populating the table shell in the Appendix. Categorical variables will be presented as counts (percent) and continuous variables will be presented as location (spread) as appropriate. Comparisons will be made with t-tests or Mann-Whitney tests as appropriate for continuous variables and chi-squared or Fisher’s exact tests as appropriate for categorical variables. No adjustment for multiple testing will be made as these are descriptive comparisons and not part of our primary hypotheses.

## Analysis Plan for Aim 1

We want to compare the mean serum K between those with CKD and those without CKD. We assume that these two populations are independent and normally distributed with unknown mean and unequal variances. We will assume that serum K among CKD positive individuals ~ and that serum K among CKD negative individuals ~ .

The two methods we will use to determine if the variances are equal or not are the variance ratio test and the rule of thumb. For the rule of thumb, we will see if the ratio of the variances is greater than or equal to 3 and use the Satterthwaite approximation for the degrees of freedom if the variances are not equal.

We will use a two-sample t-test for the hypothesis test. If the variances are equal, we will pool the variances and use degrees of freedom . If the variances are not equal, we will use each variance individually in the formulas and use the Satterthwaite approximation to find the degrees of freedom for the test. Our hypotheses are where is the unknown population mean serum K among CKD positive individuals and is the unknown population mean serum K among CKD negative individuals.

We also want to test the association between serum K and CKD status. We will use the chi-squared test to test this association. We will use a chi-squared test if the sample sizes are large enough (the count of each cell in the contingency table is greater than 5). If the sample sizes are not large enough, we will conduct a Fisher’s exact test.

The hypotheses we will test with the chi-squared test are

serum K and CKD status are independent of each other.

serum K and CKD status are associated with each other.

After conducting the chi-squared test, we will measure relative risk and the odds ratio. In this case, the relative risk represents the increase or decrease in risk of having chronic kidney disease (CKD) given an individual has normal serum K intake rather than low-normal serum K intake. The odds ratio represents the increase or decrease in odds of having CKD given an individual has normal serum K intake rather than low-normal serum K intake. We will also give the corresponding 95% confidence intervals for each of these measures. In our calculations, will be the probability that an individual is CKD positive given that they have normal serum K intake and will be the probability that an individual is CKD positive given that they have low-normal serum K intake.

Based on our research hypothesis, we expect the measures (RR and OR) to be less than 1. We hypothesize that a higher serum K will lead to negative CKD status and an odds/risk ratio greater than 1 would mean that a higher serum K will lead to positive CKD status.

Lastly, we will perform a Mantel-Haenszel (MH) test to assess the association between serum K and CKD after adjusting for age. We believe age will be a confounding variable in this analysis, so we adjust for it in this test. We will put individuals into two categories for age: 60 and older, under 60 years old. As part of the results of the MH test, we will estimate the odds ratio for each age group.

Our hypotheses for the MH test are similar to the chi-squared test:

serum K and CKD status are independent of each other after controlling for age.

serum K and CKD status are associated with each other after controlling for age.

## Analysis Plan for Aim 2

Similar analyses described in Section 4.1 above will be performed with serum K replaced with dietary K intake.

## Analysis Plan for Aim 3

We want to compare the mean eGFR between those with low-normal serum K and those with normal serum K. We assume that these two populations are independent and normally distributed with unknown mean and unequal variances. We will assume that eGFR among individuals with low-normal serum K ~ and that eGFR among individuals with normal serum K ~ .

The two methods we will use to determine if the variances are equal or not are the variance ratio test and the rule of thumb. For the rule of thumb, we will see if the ratio of the variances is greater than or equal to 3 and use the Satterthwaite approximation for the degrees of freedom if the variances are not equal.

We will use a two-sample t-test for the hypothesis test. If the variances are equal, we will pool the variances and use degrees of freedom . If the variances are not equal, we will use each variance individually in the formulas and use the Satterthwaite approximation to find the degrees of freedom for the test. Our hypotheses are where is the unknown population mean eGFR among individuals with low-normal serum K and is the unknown population mean eGFR among individuals with normal serum K.

## Analysis Plan for Aim 4

Similar analyses described in Section 4.3 above will be performed with serum K replaced with dietary K intake.

# Statistical Analysis Results

### Exploratory Analyses and CONSORT Numbers

A few outliers were noted and changed to be missing values in the following variables: BMI, SBP, ACR, A1C, and DBP.

CONSORT Numbers: There were 47 participants missing CKD status and 3 participants missing serum K values. These 50 participants were excluded from the analysis and thus the analysis was performed on 300 individuals. In addition, there were 32 participants with missing dietary K values. Thus, any analysis where dietary K is involved only includes 268 individuals.

Combined borderline adequate and adequate intake groups of diet K to make one adequate group that can be compared to the inadequate diet K intake group.

**eGFR**

The mean of eGFR is 89.34 and the standard deviation of eGFR is 23.80. Using these numbers, the interval of values of eGFR within one standard deviation of the mean is (65.54, 113.14). Exactly 68% of the eGFR values are contained in this interval. The interval of eGFR values within two standard deviations of the mean is (41.75, 136.93). 97% of the eGFR values are contained in this interval, which is close to the 95% that would be predicted assuming eGFR is approximately normally distributed.

We can construct a 95% confidence interval for eGFR within the two sub-populations of serum K using the t-distribution since we assume eGFR is normally distributed but do not know the population variance. For low-normal serum K, the 95% CI is (88.54, 96.06). We are 95% confident that the true mean for eGFR in patients with low-normal serum K levels is between 88.54 and 96.06. For normal serum K, the 95% CI is (83.02, 90.69). We are 95% confident that the true mean for eGFR in patients with normal serum K levels is between 83.02 and 90.69.

We can also look at the 95% confidence interval for the sub-populations of diet K for eGFR. For inadequate diet K intake, the 95% CI is (86.34, 93.08). Thus, we are 95% confident that the true mean for eGFR in patients with inadequate diet K intake is between 86.34 and 93.08. For borderline adequate diet K intake, the 95% CI is (82.85, 92.11). We are 95% confident that the true mean for eGFR in patients with borderline adequate diet K intake is between 82.85 and 92.11. Finally, the 95% CI for participants with adequate diet K intake is (28.95, 148.44). We are 95% confident that the true mean for eGFR in patients with adequate diet K intake is between 28.95 and 148.44. This confidence interval is very wide because there are only 3 participants with adequate diet K intake. So, we combine borderline adequate and adequate diet K intake into one group. The new 95% CI for this group is (83.00, 92.08). We are 95% confident that the true mean for eGFR in patients with borderline adequate or adequate diet K intake is between 83.00 and 92.08.

### Graphical Displays

Chart, box and whisker chart

Description automatically generatedAge: The two graphs below show that age is not normally distributed because there are participants from all age groups in the study. If age were normally distributed, there would be a lot of 40-60 year old participants and few young and old participants.

Chart, histogram

Description automatically generated

BMI: The graphs below show that BMI is not normally distributed but is actually right skewed. Most participants have a BMI between 25 and 35, but there are various participants with a BMI of over 40, even up to 70.

Chart, box and whisker chart

Description automatically generatedChart, histogram

Description automatically generated

Chart, histogram

Description automatically generatedChart, box and whisker chart

Description automatically generatedSerum K: The graphs below show that serum K is approximately normally distributed because it has the bell shaped curve with the most values between 3.75 and 4.25 and the frequency decreases as we get to the serum K values farther from the mean.

Chart, histogram

Description automatically generatedDiet K: The two graphs below show that Diet K is almost approximately normal. The few outliers greater than 2500 make it difficult to state that Diet K is approximately normal. Disregarding those outliers, the variable looks to be approximately normally distributed.

Chart, box and whisker chart

Description automatically generated

eGFR: The two graphs below show that eGFR is approximately normally distributed. The variable does not have the perfect bell-shaped curve, but it has a peak in the middle around 90-110 and the frequency decreases as it approaches the tails.

Chart, histogram

Description automatically generatedChart, box and whisker chart

Description automatically generated

Below we see the normal QQ plot and histogram for serum K in both those without CKD and those with CKD to check that they are normally distributed.

Chart, histogram

Description automatically generatedChart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, histogram

Description automatically generated

Below we see the normal QQ plot and histogram for eGFR in both those low-normal serum K and those with normal serum K.

Chart, histogram

Description automatically generatedChart, scatter chart

Description automatically generated

### Chart, histogram Description automatically generatedChart, scatter chart Description automatically generated

### Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Overall | Low-Normal K  (N=137) | Normal K  (N=163) | P |
| Age (years) | 52 (29) | 47 (28) | 55 (29) | 0.047\* |
| Sex |  |  |  | 0.023 |
| Female | 145 (48) | 76 (55) | 69 (42) |  |
| Male | 155 (52) | 61 (45) | 94 (58) |  |
| Income |  |  |  | 0.342 |
| Under $20,000 | 56 (20) | 20 (16) | 36 (24) |  |
| $20,000 to <$45,000 | 75 (27) | 32 (26) | 43 (28) |  |
| $45,000 to <$75,000 | 58 (21) | 28 (22) | 30 (20) |  |
| $75,000 and over | 78 (28) | 41 (33) | 37 (24) |  |
| Refused/don’t know | 11 (4) | 3 (3) | 7 (5) |  |
| Diet K (mg/1000kcal) | 1269 (394) | 1252 (399) | 1282 (391) | 0.538 |
| Diet K intake adequacy |  |  |  | 0.822 |
| Inadequate | 207 (77) | 95 (78) | 112 (77) |  |
| Borderline |  |  |  |  |
| Adequate | 61 (23) | 27 (22) | 34 (23) |  |

## Analysis Results for Aim 1

In our sample, there were 64 individuals that had CKD and 236 individuals that did not have CKD. We compared the variances using both the rule of thumb and the variance ratio test.

For the rule of thumb, we found that the ratio of the variances is 1.365 which is not greater than 3.

For the variance ratio test, we did a hypothesis test with the following steps:

Step 1: We assumed that serum K among CKD positive and CKD negative individuals are independent and normally distributed.

Step 2: We set up our hypotheses: where is the population variance of serum K among CKD positive individuals and is the population variance of serum K among CKD negative individuals.

Step 3: Under ,

Step 4: F = The p-value is 2 x

Step 5: Using an alpha level of 0.05, we concluded that there is not enough evidence to say that the variances are different from each other.

Thus, we chose to assume equal variances for the two-sample t-test. Since we assume equal variances for the two-sample t-test, we pooled the variances for the t-test using the following formula:

Now that we had the pooled variance and knew the correct two-sample t-test formula to use, we ran our hypothesis test to compare mean serum K between individuals with and without CKD using the following steps:

Step 1: We assumed that the two populations are independent, that serum K among CKD positive individuals ~ and that serum K among CKD negative individuals ~ .

Step 2: We set up our hypotheses: where is the unknown difference in mean serum K between CKD positive individuals and CKD negative individuals.

Step 3: Under

Step 4: T =

The p-value is 2 x P(T > 0.419) = 0.675

Step 5: Using an alpha level of 0.05, we concluded that there is not sufficient evidence to say that mean serum K is different between CKD positive and CKD negative individuals.

In the context of the problem, the t-test results mean that there is no association between CKD status and serum K. We hypothesized that higher serum K would be associated with CKD negative status, however, there was no significant difference in mean serum K between CKD positive and CKD negative individuals, so the results do not align with our hypothesis.

The 95% confidence interval for this aim was (-0.081, 0.125). We are 95% confident that the true mean difference in serum K between CKD positive and CKD negative individuals is between

-0.081 and 0.125.

Next, we conducted a chi-squared test to see if there is an association between serum K and CKD. We chose a chi-squared test because the sample size in each group (normal serum K and CKD positive, low-normal serum K and CKD positive, etc.) was greater than 5.

The results of the chi-squared test were that and the p-value = 0.529. So, we failed to reject the null hypothesis and concluded that there is no evidence against serum K and CKD being independent of each other.

The relative risk was 1.152 with a 95% CI of (0.741, 1.790). This means that individuals with normal serum K have 15.2% increased risk of having CKD compared to individuals with low-normal serum K, though this increase was not statistically significant. We are 95% confident that the true relative risk is between 0.741 and 1.790. This means that we are 95% confident that the risk for individuals with normal serum K to have CKD is between 25.9% lower than individuals with low-normal serum K and 79% higher than individuals with low-normal serum K.

The odds ratio was 1.196 with a 95% CI of (0.685, 2.091). This means that individuals with normal serum K have 19.6% higher odds of having CKD compared to individuals with low-normal serum K, though this increase was not statistically significant. We are 95% confident that the true odds ratio is between 0.685 and 2.091. This means that we are 95% confident that the odds of having CKD are somewhere between 31.5% lower and 109.1% higher for individuals with normal serum K compared to individuals with low-normal serum K.

Looking back at our research hypothesis, the relative risk and odds ratio were not in the expected direction. Instead, higher serum K is associated with greater risk and odds of having CKD.

The last part of the analysis investigated whether serum K and CKD are associated with each other after adjusting for age. There were 109 participants aged 60 and over and 191 participants age under 60.

For the participants aged 60 and over, the odds ratio was 1.138 with a 95% CI of (0.511, 2.534). This means that individuals aged 60 and over with normal serum K have 13.8% higher odds of having CKD compared to individuals aged 60 and over with low-normal serum K, though this increase was not statistically significant. We are 95% confident that the true odds ratio is between 0.511 and 2.534. This means that we are 95% confident that the odds of having CKD are somewhere between 48.9% lower and 153.4% higher for individuals aged 60 and over with normal serum K compared to individuals aged 60 and over with low-normal serum K.

For the participants aged under 60, the odds ratio was 0.75 with a 95% CI of (0.300, 1.873). This means that individuals aged under 60 with normal serum K have 25% lower odds of having CKD compared to individuals aged under 60 with low-normal serum K, though this increase was not statistically significant. We are 95% confident that the true odds ratio is between 0.300 and 1.873. This means that we are 95% confident that the odds of having CKD are somewhere between 70% lower and 87.3% higher for individuals aged under 60 with normal serum K compared to individuals aged under 60 with low-normal serum K.

These odds ratios are intriguing because there is a difference in the ratio after adjusting for age. The odds ratio for participants aged 60 and over is similar to the odds ratio where we don’t account for age. This odds ratio does not go in the expected direction as we hypothesized in the beginning. However, the odds ratio for participants aged under 60 does agree with our hypothesis.

From the Mantel-Haenszel test, we found the MH and the p-value = 0.866 after taking out the continuity correction. Thus, we failed to reject the null hypothesis and concluded that there is no evidence against serum K and CKD being independent of each other after controlling for age. This is the same conclusion as we had with the unadjusted results from the chi-squared test, but it seems even more clear now that serum K and CKD status are not associated.

## Analysis Results for Aim 2

In our sample, there were 64 individuals that had CKD and 236 individuals that did not have CKD. We compared the variances using both the rule of thumb and the variance ratio test.

For the rule of thumb, we found that the ratio of the variances is 1.095 which is not greater than 3.

For the variance ratio test, we did a hypothesis test with the following steps:

Step 1: We assumed that dietary K among CKD positive and CKD negative individuals are independent and normally distributed.

Step 2: We set up our hypotheses: where is the population variance of dietary K among CKD positive individuals and is the population variance of dietary K among CKD negative individuals.

Step 3: Under ,

Step 4: F = The p-value is 2 x

Step 5: Using an alpha level of 0.05, we concluded that there is not enough evidence to say that the variances are different from each other.

Thus, we chose to assume equal variances for the two-sample t-test. Since we assume equal variances for the two-sample t-test, we pooled the variances for the t-test using the following formula:

Now that we had the pooled variance and knew the correct two-sample t-test formula to use, we ran our hypothesis test to compare mean dietary K between individuals with and without CKD using the following steps:

Step 1: We assumed that the two populations are independent, that dietary K among CKD positive individuals ~ and that dietary K among CKD negative individuals ~ .

Step 2: We set up our hypotheses: where is the unknown difference in mean dietary K between CKD positive individuals and CKD negative individuals.

Step 3: Under

Step 4: T =

The p-value is 2 x P(T < -0.135) = 0.893

Step 5: Using an alpha level of 0.05, we concluded that there is not sufficient evidence to say that mean dietary K is different between CKD positive and CKD negative individuals.

In the context of the problem, the t-test results mean that there is no association between CKD status and dietary K. We hypothesized that higher dietary K would be associated with CKD negative status, however, there was no significant difference in mean dietary K between CKD positive and CKD negative individuals, so the results do not align with our hypothesis.

The 95% confidence interval for this aim was (-117.00, 102.02). We are 95% confident that the true mean difference in dietary K between CKD positive and CKD negative individuals is between

-117.00 and 102.02.

Next, we conducted a chi-squared test to see if there is an association between dietary K and CKD. We chose a chi-squared test because the sample size in each group (inadequate dietary K and CKD positive, borderline adequate or adequate dietary K and CKD positive, etc.) was greater than 5.

The results of the chi-squared test were that and the p-value = 0.928. So, we failed to reject the null hypothesis and concluded that there is no evidence against dietary K and CKD being independent of each other.

The relative risk was 1.026 with a 95% CI of (0.591, 1.779). This means that individuals with borderline adequate or adequate dietary K have 2.6% increased risk of having CKD compared to individuals with inadequate dietary K, though this increase was not statistically significant. We are 95% confident that the true relative risk is between 0.591 and 1.779. This means that we are 95% confident that the risk for individuals with borderline adequate or adequate dietary K to have CKD is between 40.9% lower than individuals with inadequate dietary K and 77.9% higher than individuals with inadequate dietary K.

The odds ratio was 1.033 with a 95% CI of (0.514, 2.078). This means that individuals with borderline adequate or adequate dietary K have 3.3% higher odds of having CKD compared to individuals with inadequate dietary K, though this increase was not statistically significant. We are 95% confident that the true odds ratio is between 0.514 and 2.078. This means that we are 95% confident that the odds of having CKD are somewhere between 48.6% lower and 107.8% higher for individuals with borderline adequate or adequate dietary K compared to individuals with inadequate dietary K.

Looking back at our research hypothesis, the relative risk and odds ratio were not in the expected direction. Instead, higher dietary K is associated with greater risk and odds of having CKD.

The last part of the analysis investigated whether dietary K and CKD are associated with each other after adjusting for age. There were 109 participants aged 60 and over and 191 participants age under 60.

For the participants aged 60 and over, the odds ratio was 1.146 with a 95% CI of (0.464, 2.829). This means that individuals aged 60 and over with borderline adequate or adequate dietary K have 14.6% higher odds of having CKD compared to individuals aged 60 and over with inadequate dietary K, though this increase was not statistically significant. We are 95% confident that the true odds ratio is between 0.464 and 2.829. This means that we are 95% confident that the odds of having CKD are somewhere between 53.6% lower and 182.9% higher for individuals aged 60 and over with borderline adequate or adequate dietary K compared to individuals aged 60 and over with inadequate dietary K.

For the participants aged under 60, the odds ratio was 0.465 with a 95% CI of (0.102, 2.127). This means that individuals aged under 60 with borderline adequate or adequate dietary K have 53.5% lower odds of having CKD compared to individuals aged under 60 with inadequate dietary K, though this increase was not statistically significant. We are 95% confident that the true odds ratio is between 0.102 and 2.127. This means that we are 95% confident that the odds of having CKD are somewhere between 89.8% lower and 112.7% higher for individuals aged under 60 with borderline adequate or adequate dietary K compared to individuals aged under 60 with inadequate dietary K.

These odds ratios are intriguing because there is a difference in the ratio after adjusting for age. The odds ratio for participants aged 60 and over is closer to the odds ratio where we don’t account for age. This odds ratio does not go in the expected direction as we hypothesized in the beginning. However, the odds ratio for participants aged under 60 does agree with our hypothesis.

From the Mantel-Haenszel test, we found the MH and the p-value = 0.715 after taking out the continuity correction. Thus, we failed to reject the null hypothesis and concluded that there is no evidence against dietary K and CKD being independent of each other after controlling for age. This is the same conclusion as we had with the unadjusted results from the chi-squared test, but it seems even more clear now that dietary K and CKD status are not associated.

## Analysis Results for Aim 3

In our sample, we had 137 individuals with low-normal serum K and 163 individuals with normal serum K. We compared the variances using both the rule of thumb and the variance ratio test.

For the rule of thumb, we found that the ratio of the variances is 1.242 which is not greater than 3.

For the variance ratio test, we did a hypothesis test with the following steps:

Step 1: We assumed that eGFR among individuals with low-normal serum K and individuals with normal serum K are independent and normally distributed.

Step 2: We set up our hypotheses: where is the population variance of eGFR among individuals with low-normal serum K and is the population variance of eGFR among individuals with normal serum K.

Step 3: Under ,

Step 4: F = The p-value is 2 x

Step 5: Using an alpha level of 0.05, we concluded that there is not enough evidence to say that the variances are different from each other.

Thus, we chose to assume equal variances for the two-sample t-test. Since we assume equal variances for the two-sample t-test, we pooled the variances for the t-test using the following formula:

Now that we had the pooled variance and knew the correct two-sample t-test formula to use, we ran our hypothesis test to compare mean eGFR between individuals with low-normal serum K and individuals with normal serum K using the following steps:

Step 1: We assumed that the two populations are independent, that eGFR among individuals with normal serum K ~ and that eGFR among individuals with low-normal serum K ~ .

Step 2: We set up our hypotheses: where is the unknown difference in mean eGFR between individuals with normal serum K and individuals with low-normal serum K.

Step 3: Under

Step 4: T =

The p-value is 2 x P(T < -1.984) = 0.048

Step 5: Using an alpha level of 0.05, we concluded that there is sufficient evidence to say that mean eGFR is different between individuals with low-normal serum K and individuals with normal serum K.

In the context of the problem, the t-test results mean that there is an association between serum K and eGFR. We hypothesized that higher serum K will be associated with better kidney function. Our hypothesis is correct that there is an association between serum K and kidney function. However, since eGFR is higher for individuals with low-normal serum K and higher eGFR is associated with better kidney function, the results do not align with our hypothesis. Lower serum K is associated with better kidney function.

The 95% confidence interval for this aim was (-10.847, -0.044). We are 95% confident that the true mean difference in eGFR between individuals with normal serum K and individuals with low-normal serum K is between -10.847 and -0.044.

## Analysis Results for Aim 4

In our sample, we had 207 individuals that had inadequate dietary K intake and 61 individuals that had borderline adequate or adequate dietary K intake. We compared the variances using both the rule of thumb and the variance ratio test.

For the rule of thumb, we found that the ratio of the variances is 1.925 which is not greater than 3.

For the variance ratio test, we did a hypothesis test with the following steps:

Step 1: We assumed that eGFR among individuals with inadequate dietary K intake and individuals with borderline adequate or adequate dietary K intake are independent and normally distributed.

Step 2: We set up our hypotheses: where is the population variance of eGFR among individuals with inadequate dietary K intake and is the population variance of eGFR among individuals with borderline adequate or adequate dietary K intake.

Step 3: Under ,

Step 4: F = The p-value is 2 x

Step 5: Using an alpha level of 0.05, we concluded that there is enough evidence to say that the variances are different from each other.

Although the result of the rule of thumb indicated that equal variances could be used, we chose to assume unequal variances for the two-sample t-test based on the result of the variance ratio test. Since we assumed unequal variances for the two-sample t-test, we used the Satterthwaite approximation to find the degrees of freedom for the t-test using the following formula:

So

Now that we had our new degrees of freedom and knew the correct two-sample t-test formula to use, we ran our hypothesis test to compare mean eGFR between individuals with inadequate dietary K intake and individuals with borderline adequate or adequate dietary K intake using the following steps:

Step 1: We assumed that the two populations are independent, that eGFR among individuals with inadequate dietary K intake ~ and that eGFR among individuals with borderline adequate or adequate dietary K intake ~ .

Step 2: We set up our hypotheses: where is the unknown difference in mean eGFR between individuals with inadequate dietary K intake and individuals with borderline adequate or adequate dietary K intake.

Step 3: Under

Step 4: T =

The p-value is 2 x P(T > 0.764) = 0.446

Step 5: Using an alpha level of 0.05, we concluded that there is not sufficient evidence to say that mean eGFR is different between individuals with inadequate dietary K intake and individuals with borderline adequate or adequate dietary K intake.

In the context of the problem, the t-test results mean that there is no association between dietary K intake and eGFR. We hypothesized that higher dietary K will be associated with better kidney function, however, there was no significant difference in mean eGFR between individuals with inadequate dietary K intake and individuals with borderline or adequate dietary K intake, so the results do not align with the hypothesis.

The 95% confidence interval for this aim was (-3.450, 7.791). We are 95% confident that the true mean difference in eGFR between individuals with inadequate dietary K intake and individuals with borderline or adequate dietary K intake is between -3.450 and 7.791.

# Addendum for Post-Hoc Analyses

Not needed for this analysis.

# Appendix

Table 1: Demographic and clinical characteristics of study participants at

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Overall | Low-Normal K  (N= ) | Normal K  (N= ) | P |
| Age (years) | location (spread) | location (spread) | location (spread) | XX |
| Sex |  |  |  | XX |
| Female | n (%) | n (%) | n (%) |  |
| Male | n (%) | n (%) | n (%) |  |
| Income |  |  |  | XX |
| Under $20,000 | n (%) | n (%) | n (%) |  |
| $20,000 to <$45,000 | n (%) | n (%) | n (%) |  |
| $45,000 to <$75,000 | n (%) | n (%) | n (%) |  |
| $75,000 and over | n (%) | n (%) | n (%) |  |
| Refused/don’t know | n (%) | n (%) | n (%) |  |
| Diet K (mg/1000kcal) | location (spread) | location (spread) | location (spread) | XX |
| Diet K intake adequacy |  |  |  |  |
| Inadequate | n (%) | n (%) | n (%) |  |
| Borderline |  |  |  |  |
| Adequate | n (%) | n (%) | n (%) |  |

# References

None.