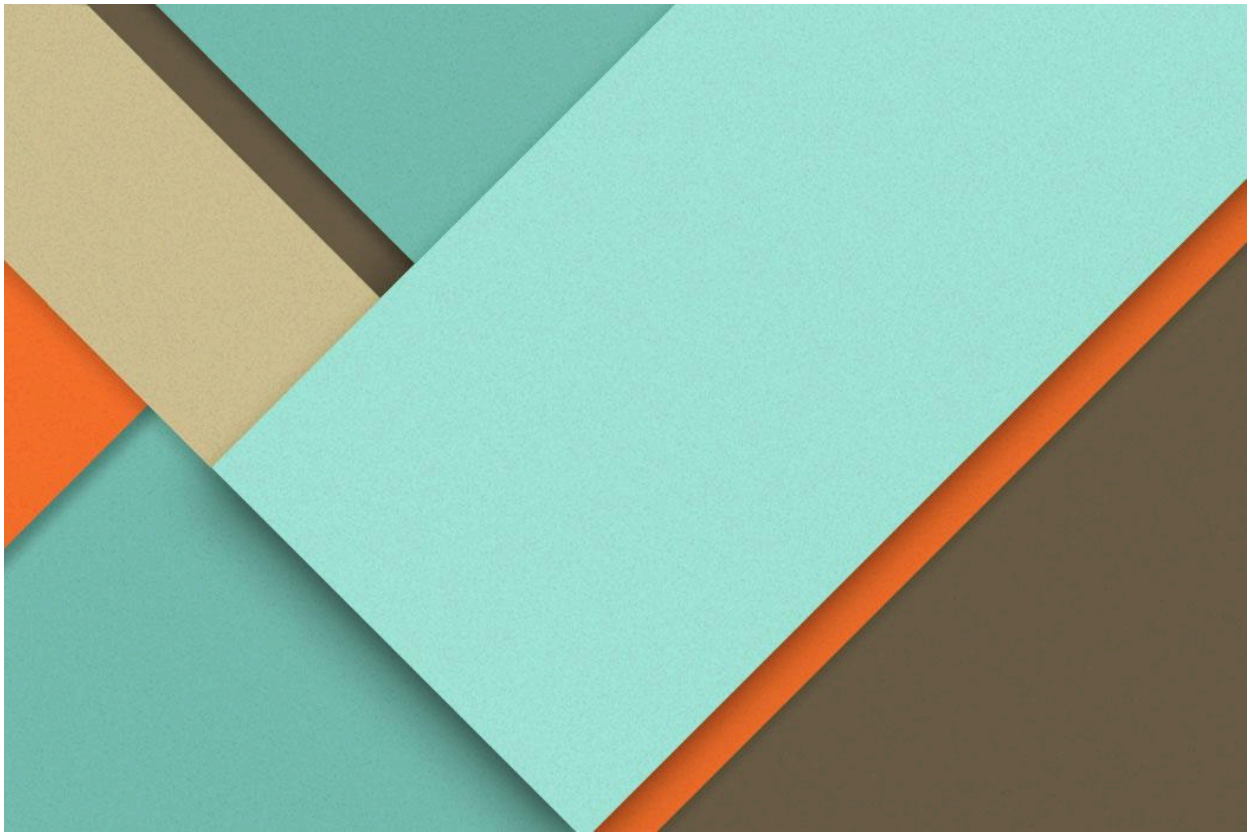


CKD Screening Tool

02.24.2019

—

Group Nine

Alya Nadji, Will Crowley, Trevor John, Brian Weiss, Brandon Fitts, Heng Zhang

STAT 630 - 674

Introduction

This report describes the development of a predictive model to identify patients at increased risk for chronic kidney disease (CKD) and discusses the construction of an effective screening tool to help identify those at risk of the illness. CKD is a progressive condition that can have serious impacts on quality of life. Latent CKD can lead to end-stage renal disease and ultimately kidney failure, requiring dialysis or a kidney transplant. These treatments are costly and increase exponentially with latent stages of CKD[1] ($12,700 per person per year with stage 4). In the United States CKD affects roughly 14% of the overall population, however less than 10% of patients with stage 1-3 CKD know they have the disease (see Appendix A) (National Institute of Diabetes and Digestive and Kidney Diseases, 2016). Undetected CKD leads to greater overall healthcare costs and increased rates of morbidity and mortality as patients with undetected CKD advance through to end-stage renal disease. Screening entire populations for CKD is a costly undertaking. Therefore, it is a priority in the field of public health to identify a targeted population for CKD screening, to reduce costs and target those at highest risk.

For each patient correctly identified as having CKD (“true positive”), we were rewarded $1,300; each patient incorrectly identified as having CKD (“false positive”) costs us $100. The predictive model and our chosen factors was utilized to create our survey instrument. We then used our predictive model to build our screening tool. Our aim is to develop an easy to use tool that can be interpreted by someone without statistical training.

Background

The CKD case study[2] contained 33 demographic (age, sex, etc) and health factors (cholesterol, hypertension, etc) from 8,819 adults aged 20 to 85 (see Appendix B). This case study population is a non-random sample of U.S. adults and therefore our predictions cannot be extrapolated to any population systematically different than those in our training set.

In addition, we reviewed relevant literature to provide context for CKD health outcomes and to inform our prediction methodology. One of the key pieces of research was a critical assessment of multiple CKD risk models[3] completed in 2012 from the National Center for Biotechnology Information (NCBI). This report analyzed not only the methodologies of

[1] "Medical Costs of CKD in the Medicare Population | American ... - JASN." https://jasn.asnjournals.org/content/24/9/1478. Accessed 26 Feb. 2019.
[2] "Screening for Chronic Kidney Disease - Darden Business Publishing." 11 Jul. 2007, http://store.darden.virginia.edu/screening-for-chronic-kidney-disease. Accessed 26 Feb. 2019.
[3] "Risk Models to Predict Chronic Kidney Disease and Its ... - NCBI - NIH." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3502517/. Accessed 26 Feb. 2019.

various CKD prediction models but also their relative effectiveness. Additionally, we reviewed other CKD prediction assessment tools to get a better understanding of risk factors associated with CKD (age, race, blood pressure). A good example of this would be the CKD Risk Map[4] created by the National Kidney Foundation. Lastly, we looked at already existing screening tools and questionnaires for CKD prediction. We then considered this foundation in the literature to pave decision-making for our analysis. Both the questions included and how they were asked (categorical, rating, etc) helped to inform our the creation of our final patient screening tool.

Methods

Given the outcome of CKD is dichotomous, we decided to utilize logistic regression as our key method for our predictive model. We ran simple summary statistics to evaluate any major differences among our patients with a known CKD status and those without and ran a correlation matrix. Initially, some strong correlations became apparent, SBP is highly correlated with hypertension; stroke is highly correlated with CVD. We noticed instances of multicollinearity, which would help to inform some of the factors we removed (total cholesterol and LDL).

We minimized and modified variables in our dataset to reduce 33 variables down to 21 variables for many reasons (see appendix E). These efforts improve our confidence in our model. For example, we removed some variables that measured similar phenomena, through collinearity. Variables were removed if collinearity was present in our correlation matrix (see Appendix F). Another example is with age which is highlighted for its importance later in our report.

We did not standardize our data to improve interpretability for non-statistical personnel (average patient). Given the multiple stakeholders-patients, providers, family members, etc.-involved in a project like this, we wanted to keep the varying scales and units for each of the variables that were relevant to our model. For example, standardized weight and blood pressure variables could be very difficult to interpret for those with limited statistical knowledge.

Another key step of logistic regression and building our model was determining the probability threshold at which someone has an increased chance of developing CKD based on our model. We based our initial threshold in our review of other screening tools currently in use and summary reports. Through some trial and error, we eventually settled on a threshold of 30% as it provided the best balance between cost reduction and not having an outsized number of false positives and false negatives. We decided to be more

[4] "Quick Reference Guide on Kidney Disease Screening | National …." https://www.kidney.org/kidneydisease/siemens_hcp_quickreference. Accessed 26 Feb. 2019.

inclusive of patients that may not have CKD (false positives), than to focus more on financial return.

With our 21 variables, we leveraged stepwise regression through backward elimination on only patients with complete values across our variables of interest. As such, variables with p-values > 0.15 were removed. We then chose to utilize our best performing model, based upon a combination of accuracy and costs. After some trial and error, we identified a set of variables that revealed statistically significant relationships based on their probabilities and potential contribution towards someone having the disease. We verified our updated, refined model and the variables we included based on looking at their confidence intervals.

Our final model included the following variables and their odds interpretation (reasonings for this modification are explained later):

- Sex: Someone who answers that they are female has about 1.5 times the odds of having CKD than someone who answers they are not female, holding all other variables constant.
- Hispanic: Someone who answers that they are hispanic have about 0.5 times the odds of having CKD than someone who answers they are not hispanic, holding all other variables constant.
- Age: Someone who answers that they are 0-40, 41-50, 51-60, 61-74, and 75+ have about 0, 3, 5.5, 16, and 53 times the odds of having CKD than someone who answers they are not in that age group, holding all other variables constant.
- Hypertension: Someone who answers that they have hypertension have about 2 times the odds of having CKD than someone who answers they do not have hypertension, holding all other variables constant.
- Anemia: Someone who answers that they have anemia have about 3 times the odds of having CKD than someone who answers they do not have anemia, holding all other variables constant.
- CVD: Someone who answers that they have CVD have about 2 times the odds of having CKD than someone who answers they do not have CVD, holding all other variables constant.
- HDL: Someone who answers that they are female has about 1.5 times the odds of having CKD than someone who answers they are not female, holding all other variables constant.
- DBP: Someone who answers that they have high blood pressure has about 1 times the odds of having CKD than someone who answers they do not have high blood pressure, holding all other variables constant.
- HDL: Someone who answers that they have high cholesterol have about 1 times the odds of having CKD than someone who answers they do not have high cholesterol, holding all other variables constant.

- PVD: Someone who answers that they have PVD have about 1.5 times the odds of having CKD than someone who answers they do not have PVD, holding all other variables constant.
- Diabetes: Someone who answers that they have diabetes have about 2 times the odds of having CKD than someone who answers they do not have diabetes, holding all other variables constant.
- Activity Level: Someone who answers that they are very active have about 1 times the odds of having CKD than someone who answers they are not very active, holding all other variables constant.

We can see that age is an extreme predictor in our model for CKD which is supported by the literature. After developing a finalized model based upon the patients with complete data in our model's variables patients in the training dataset, we performed imputations on the 2819 patients with no disease status. We leveraged the DMwR package in R for k-nearest neighbor imputation, where k=3. For a given patient with missing values, we compared their attributes across the dataset with 3 other similar patients and replaced missing values with averages of those nearest neighbor patients. Given that our model greatly reduced the dataset, we relied less on imputation techniques for our prediction.

Prediction Results

For predicting CKD status for the 2819 patients without a status, we relied on our imputation technique for about 6% of the NA values in the 2819 patients for overall CKD prediction (see Appendix D).

The final model on the training set identified 92.19% of patients with true positive and true negative disease statuses (our accuracy). However, our model incorrectly identified 197 false positives (type I error) patients and 178 false negative (type II error) participants. The entire outputs and results of our model can be found in Appendix G. The resulting financial implications of our model were:

- Total Revenue: $205,400
- Total Costs Incurred: ($55,300)
- Total Returns: $150,100

We ran our model on the 2,819 sample group and predicted a 9% CKD disease prevalence. While this is higher than the overall prevalence in the entire dataset, it is reasonably close to 7% and also allows us to catch more that do not have the disease and we say that they do (false positive), instead of missing more that do have the disease that we did not catch (false negatives).

After discussing our model and its predictions, we designed a simple and transparent screening tool. Our questions are based on our final variables in our model and use the coefficients to score our instrument. We transformed the log odds to odds to determine

the weights for each variable, and our scores for each answer are based on these weights: firstly, we transform log odds to odds ($e^{\beta 1}$) = The odds of having CKD in relation to X_1 (see Appendix H). For example, the odds of Anemia was 2.77, rounded up to 3 for simplicity and ease of calculation for screening taker; therefore, if a person answered yes to our question "Do you have anemia or have you been told by a doctor that you have anemia?", they would be assigned 3 points. The wording of each screening question was very deliberate given the previously stated focus on simplicity and ease of use. Ultimately, an individual, based on their age, can evaluate if they are at an increased chance for CKD, and therefore whether they should consult their doctor about further CKD testing.

In order to determine the compatibility of our predictive model probabilities and our screening tool totals, we calculated each individuals totals from our training set as if they had filled our our screening tools. We then ran a correlation between the computed odds or each persons screening tool point totals and the associated probabilities given by our prediction model for the training set patients. Our screening tool scoring system appropriately reflects our fitted model with a strong correlation of 0.87 the two variables.

Limitations

This analysis has five concerning limitations.

First, some of the main CKD risk factors we came across in our literature research such as family history of kidney disease, serum creatinine, frequent NSAID usage, and others, were not included in our dataset.

Second, by prioritizing ease of use for our screening tool, we removed some potential factors that were represented in our predictive model.

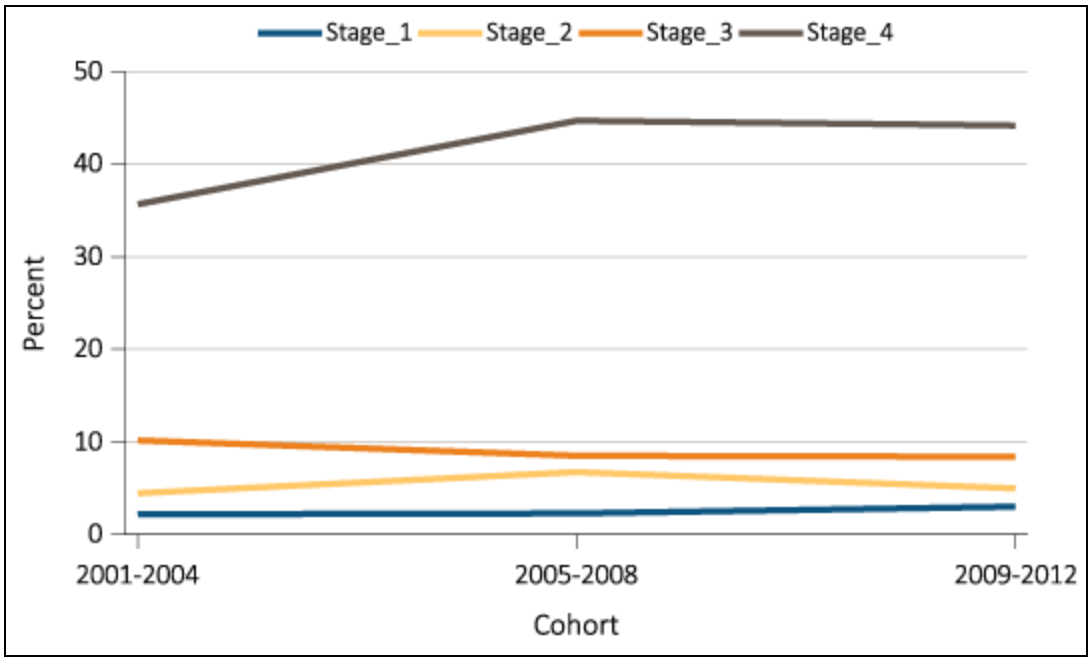
Third, we had no control over the population that comprised our training and test data; all we knew was that it was not a true representation of the global population. Therefore, external validity for our results is low and it would not be valid for certain geographies or regions, especially if their population differed greatly from our sample.

Fourth, because this was a cross sectional dataset, we are only drawing conclusions on a certain point in time. We therefore cannot determine whether or not the variables actually puts patients at a higher risk and had no insight into patterns or trends associated with some of the variables. Risk in epidemiological research implies causation. As such, our model predicts increased chances of having CKD based on a population similar to our training dataset.

Lastly, the meta analysis from NCBI we leveraged in our background research noted the varying quality of the predictive models. The report suggested that a lack of historical investment and focus on nephrology might be a factor explaining the limited success of models built.

Appendix

Appendix A - Percent of Disease Awareness based CKD Stage



Appendix B - Initial Variables in the Data Set

Col.	*Variable*	*Definition*
A	ID	Identification number
B	Age	Age (years)
C	Female	1 if female
D	Racegrp	Self-reported race/ethnic group (white, black, Hispanic, other)
E	Educ	1 if more than high school
F	Unmarried	1 if unmarried
G	Income	1 if household income is above the median
H	CareSource	Self-reported source of medical care (Dr./HMO, clinic, noplace, other)
I	Insured	1 if covered by health insurance
J	Weight	Weight (kg)
K	Height	Height (cm)
L	BMI	Body mass index (kg/m^2)
M	Obese	1 if BMI is greater than 30 kg/m^2
N	Waist	Waist circumference (cm)
O	SBP	Systolic blood pressure (max)
P	DBP	Diastolic blood pressure (min)
Q	HDL	(mg/dL) the "good" cholesterol
R	LDL	(mg/dL) the "bad" cholesterol
S	Total Chol	(mg/dL) the sum of good and bad cholesterol
T	Dyslipidemia	Too high LDL or too low HDL
U	PVD	Peripheral vascular disease reflected by reduced SBP at the leg relative to the arm
V	Activity	Mostly sit (1); stand or walk a lot (2); lift light loads or climb stairs often (3); heavy work and heavy loads (4)
W	Poor Vision	Self-reported poor vision
X	Smoker	Smoked at least 100 cigarettes
Y	Hypertension	The presence of at least one of four indicators of high blood pressure
Z	Fam Hypertension	Family history of hypertension (high blood pressure)
AA	Diabetes	Self-reported physician diagnosed or lab test result
AB	Fam Diabetes	Family history of diabetes
AC	Stroke	Self-reported response to "Has a doctor ever told you that you had a stroke?"
AD	CVD	Response to "Has a doctor ever told you that you had angina pectoris, myocardial infarction, or stroke?"
AE	Fam CVD	Family history of cardiovascular disease
AF	CHF	Self-reported response to "Has a doctor ever told you that you had congestive heart failure?"
AG	Anemia	Treatment for anemia received in past three months or hemoglobin at exam lower than 11g/dL
AH	CKD	Chronic kidney disease as indicated by measured serum creatinine

Appendix C - Factors Included in Models of Risk Prediction for Chronic Kidney Disease

Table S4: Factors included in models of risk prediction for chronic kidney disease

Author, Reference	Name of the risk model	Age	Sex/ gender	Ethnicity	Anemia/ Hemoglobin level	Estimated glomerular filtration rate	Proteinuria /albuminuria	Glucose level/ Diabetes/ history of diabetes	Blood pressure/ Hypertension	Cardiovascular disease	Smoking	Blood lipids s	Body mass index/ adiposity	Kidney stone / Uric acid	Aldosterone	Homocy
Bang et al, 2007 [1]	SCORED score	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	No	No	No	No	No	No
Kshirsagar et al, 2008 [2]	ARIC/CHS score 1	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	No	No	No	No	No	No
Kshirsagar et al, 2008 [2]	ARIC/CHS score 2	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	No	Yes	No	No	No	No
Fox et al, 2010[3]	Framingham score 1	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No
Fox et al, 2010[3]	Framingham score 2	Yes	Yes	No	No	Yes	No	Yes	Yes	No	Yes	Yes	No	No	No	No
Fox et al, 2010[3]	Framingham score 3	Yes	Yes	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes
Hippisley-Cox et al, 2010 [4]	QKidney score	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Chien, et al 2010[5]	Taiwan score 1	Yes	Yes	No	No	No	No	Yes	Yes	Yes	No	No	Yes	No	No	No
Chien, et al 2010[5]	Taiwan score 2	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No
Halbesma et al, 2011[6]	PREVEND score	Yes	Yes	No	No	Yes	Yes	No	Yes	No	No	No	No	No	No	No

Appendix D

Total NA's before Imputation	Total Population Missing CKD Status	% Relying on Imputation for CKD Status Prediction
173	2819	173/2819 = 6%

Appendix E - Variable and Reason for Removal From Analysis Data

Age (continuous)	We decided that the continuous Age variable would be misleading for age groups where disease prevalence is far higher than for lower age groups. We cut this variable in the following age categories: 20-40, 41-50, 51-60, 61-74, and 75-85. These age groups display the following disease prevalence: 0.4%, 1.7%, 4.3%, 12.4%, and 35.7%, respectively.
SBP	SBP & Stroke strongly correlated 0.61. Literature supports that those with CKD are at higher risk for stroke. Given this we have chosen to remove SBP.
Total Cholesterol	Total Cholesterol and LDL are essentially the same variable with 0.9 correlation. This is removed because it would be redundant to keep this variable in our model.
Unmarried	The literature does not describe unmarried as a risk factor for CKD. This variable was discarded to reduce spurious associations in our model.
BMI	BMI was an indicator created to generalize at the population level and not intended for individualistic predicting. In addition, we

	discussed as a group that BMI can be misleading in regards to one's health status.
Weight & Obese	Waist is strongly correlated with weight (0.87) and obese (0.67). Also given that obese is a categorical variable based from BMI, it also makes additional sense to remove redundancies in the model.
Waist & Height	Numerous studies have posed that a waist to height ratio is a strong predictor for CKD and even stronger than that of just waist and a better measure than obesity/BMI. Therefore Height & Waist were transformed into a ratio variable and then removed from the overall dataset (He et al., 2016)*.
Poor Vision	This variable is self-reported and is also not represented by a majority of the literature that we researched.
Family History of Hypertension	Because we are predicting current CKD status and already have Hypertension of patient, we removed family history of hypertension as it is further associated to CKD risk than that of the person's status of hypertension.

*He, Y., Li, F., Zhao, X., Zeng, Q. (2016). The association of chronic kidney disease and waist circumference and waist-to-height ratio in Chinese urban adults. *MEDICINE, 95(5)*, doi: 10.1097/MD.0000000000003769

Appendix F - Correlation Matrix of Dataset (after removing variables)

Appendix G - Log Odd to Odds

No	Variables	Coefficients (Log Odds)	Odds
1	Female	0.328918	1.389464
2	DBP	-0.007363	0.992664
3	HDL	-0.017887	0.982272
4	PVD	0.347834	1.415997
5	Activity	-0.268132	0.764807
6	Hypertension	0.720830	2.056139
7	Diabetes	0.522653	1.686496
8	CVD	0.644443	1.904926
9	Anemia	1.020318	2.774077
10	Race group hispa	-1.002519	0.366954
11	Age 41 to 50	1.111882	3.040074
11	Age 51 to 60	1.700906	5.478909
11	Age 61 to 74	2.793364	16.33588
11	Age 75 to 86	3.966306	52.78917

Appendix F: R Code Detailing Results and Costs

```
> a <- c_accuracy(data_in$CKD,classify)  # to run this you must run my code below first.
> round(a,4)
   recall precision  accuracy       tpr       fpr  fmeasure        tp        tn        fp        fn
   0.4702    0.4451    0.9219    0.4702    0.0441    0.4573  158.0000 4266.0000  197.0000  178.0000
> ## Step 9 - Caclculate Costs
> acc=c_accuracy(data_in$CKD,classify)
> c1=100   # penalize me  $100 for a false positive
> c2=200  #  penalize me $200 for a false negatives
> cost=acc[9]*c1+acc[10]*c2
> cost   ## my costs are $48,800  because I got hit with a ton of false negative costs
   fp
                                                                                                55300
```

Appendix H - linear regression

```
lm(formula = CKD ~ ., data = data_in)

Residuals:
     Min       1Q   Median       3Q      Max
-0.63431 -0.07150 -0.01761  0.01122  1.01977

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.326e-01  3.777e-02   3.510 0.000452 ***
Female              1.672e-02  7.535e-03   2.219 0.026519 *
Educ               -4.846e-03  7.309e-03  -0.663 0.507386
Income             -9.440e-05  7.393e-03  -0.013 0.989813
Insured            -1.998e-03  9.640e-03  -0.207 0.835839
Waist_HeightRatio  -5.817e-02  3.152e-02  -1.846 0.065013 .
DBP                -7.076e-04  3.001e-04  -2.358 0.018406 *
HDL                -9.874e-04  2.388e-04  -4.135 3.61e-05 ***
LDL                 1.229e-04  8.621e-05   1.425 0.154203
Dyslipidemia       -1.849e-02  1.160e-02  -1.594 0.110920
PVD                 6.763e-02  1.830e-02   3.696 0.000221 ***
Activity           -9.459e-03  4.241e-03  -2.231 0.025759 *
Smoker              2.267e-03  7.485e-03   0.303 0.762035
Hypertension        3.917e-02  8.298e-03   4.721 2.42e-06 ***
Diabetes            5.687e-02  1.147e-02   4.958 7.37e-07 ***
Stroke              4.904e-02  2.717e-02   1.805 0.071142 .
CVD                 6.438e-02  1.992e-02   3.232 0.001239 **
CHF                 5.349e-02  2.239e-02   2.389 0.016931 *
Anemia              6.314e-02  2.440e-02   2.588 0.009672 **
Racegrphispa       -3.162e-02  1.056e-02  -2.996 0.002750 **
Racegrpother       -2.277e-03  2.064e-02  -0.110 0.912149
Racegrpwhite        1.153e-02  9.780e-03   1.179 0.238552
CareSourceclinic   -1.028e-03  1.675e-02  -0.061 0.951078
CareSourceDrHMO    -4.891e-03  1.591e-02  -0.307 0.758523
CareSourcenoplace  -9.592e-04  1.754e-02  -0.055 0.956400
agecatAge41to50     5.388e-03  9.845e-03   0.547 0.584222
agecatAge51to60     6.948e-03  1.111e-02   0.625 0.531847
agecatAge61to74     7.131e-02  1.097e-02   6.499 8.94e-11 ***
agecatAge75to86     2.750e-01  1.400e-02  19.651  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2274 on 4770 degrees of freedom
Multiple R-squared:  0.2104,    Adjusted R-squared:  0.2058
F-statistic: 45.39 on 28 and 4770 DF,  p-value: < 2.2e-16

> summary(predict(model))
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.08288 -0.00197  0.02638  0.07001  0.09616  0.66313
```

Appendix I

- Logistic regression including ALL 21 of our variables:

```
Call:
glm(formula = CKD ~ ., family = "binomial", data = data_in)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9848  -0.3003  -0.1355  -0.0816   3.3637

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -3.700404   0.782200  -4.731 2.24e-06 ***
Female              0.297505   0.152464   1.951 0.051019 .
Educ               -0.090023   0.150467  -0.598 0.549645
Income             -0.003638   0.158344  -0.023 0.981669
Insured             0.186335   0.312299   0.597 0.550738
Waist_HeightRatio  -0.514369   0.505273  -1.018 0.308677
DBP                -0.007733   0.005193  -1.489 0.136474
HDL                -0.017676   0.004933  -3.583 0.000339 ***
LDL                 0.002596   0.001727   1.504 0.132674
Dyslipidemia       -0.259100   0.225556  -1.149 0.250674
PVD                 0.348198   0.205771   1.692 0.090616 .
Activity           -0.277060   0.102046  -2.715 0.006626 **
Smoker              0.056156   0.138733   0.405 0.685640
Hypertension        0.758875   0.166008   4.571 4.85e-06 ***
Diabetes            0.535559   0.163174   3.282 0.001030 **
Stroke              0.241130   0.308937   0.781 0.435087
CVD                 0.454645   0.239330   1.900 0.057478 .
CHF                 0.258391   0.256368   1.008 0.313507
Anemia              1.153233   0.449494   2.566 0.010299 *
Racegrphispa       -0.814149   0.244588  -3.329 0.000873 ***
Racegrpother        0.106639   0.479432   0.222 0.823982
Racegrpwhite        0.271802   0.194451   1.398 0.162175
CareSourceclinic   -0.116230   0.321186  -0.362 0.717445
CareSourceDrHMO    -0.226848   0.301532  -0.752 0.451859
CareSourcenoplace  -0.466815   0.439941  -1.061 0.288652
agecatAge41to50     1.056548   0.436727   2.419 0.015553 *
agecatAge51to60     1.611767   0.405445   3.975 7.03e-05 ***
agecatAge61to74     2.667251   0.372527   7.160 8.07e-13 ***
agecatAge75to86     3.782565   0.377248  10.027  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2434.8  on 4798  degrees of freedom
Residual deviance: 1639.7  on 4770  degrees of freedom
AIC: 1697.7

Number of Fisher Scoring iterations: 8
```

Appendix J

- GLM – stepwise backward logistic regression – FINAL MODEL USED

```
Call:
glm(formula = CKD ~ Female + DBP + HDL + PVD + Activity + Hypertension +
    Diabetes + CVD + Anemia + Racegrphispa + agecatAge41to50 +
    agecatAge51to60 + agecatAge61to74 + agecatAge75to86, family = "binomial",
    data = data_in)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0130  -0.3033  -0.1356  -0.0850   3.3338

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     -3.596145   0.571279  -6.295 3.08e-10 ***
Female           0.328918   0.141745   2.320 0.020314 *
DBP             -0.007363   0.005078  -1.450 0.147096
HDL             -0.017887   0.004650  -3.847 0.000120 ***
PVD              0.347834   0.202834   1.715 0.086369 .
Activity        -0.268132   0.100303  -2.673 0.007513 **
Hypertension     0.720830   0.163311   4.414 1.02e-05 ***
Diabetes         0.522653   0.159162   3.284 0.001024 **
CVD              0.644443   0.170928   3.770 0.000163 ***
Anemia           1.020318   0.440109   2.318 0.020431 *
Racegrphispa    -1.002519   0.189701  -5.285 1.26e-07 ***
agecatAge41to50  1.111882   0.434617   2.558 0.010518 *
agecatAge51to60  1.700906   0.401693   4.234 2.29e-05 ***
agecatAge61to74  2.793364   0.365252   7.648 2.04e-14 ***
agecatAge75to86  3.966306   0.365887  10.840  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2434.8  on 4798  degrees of freedom
Residual deviance: 1650.0  on 4784  degrees of freedom
AIC: 1680

Number of Fisher Scoring iterations: 8
```