

Machine Learning to Compare Frequent Medical Problems of African American and Caucasian Diabetic Kidney Patients

Yong-Mi Kim, PhD¹, Pranay Kathuria, MD, FACP, FASN, FNKF², Dursun Delen, PhD³

¹School of Library and Information Studies, University of Oklahoma, Tulsa, OK, USA; ²Division of Nephrology and Hypertension, Department of Medicine, School of Community Medicine, University of Oklahoma, Tulsa, OK, USA; ³Spears School of Business, Oklahoma State University, Tulsa, OK, USA

Objectives: End-stage renal disease (ESRD), which is primarily a consequence of diabetes mellitus, shows an exemplary health disparity between African American and Caucasian patients in the United States. Because diabetic chronic kidney disease (CKD) patients of these two groups show differences in their medical problems, the markers leading to ESRD are also expected to differ. The purpose of this study was, therefore, to compare their medical complications at various levels of kidney function and to identify markers that can be used to predict ESRD. **Methods:** The data of type 2 diabetic patients was obtained from the 2012 Cerner database, which totaled 1,038,499 records. The data was then filtered to include only African American and Caucasian outpatients with estimated glomerular filtration rates (eGFR), leaving 4,623 records. A priori machine learning was used to discover frequently appearing medical problems within the filtered data. CKD is defined as abnormalities of kidney structure, present for >3 months. **Results:** This study found that African Americans have much higher rates of CKD-related medical problems than Caucasians for all five stages, and prominent markers leading to ESRD were discovered only for the African American group. These markers are high glucose, high systolic blood pressure (BP), obesity, alcohol/drug use, and low hematocrit. Additionally, the roles of systolic BP and diastolic BP vary depending on the CKD stage. **Conclusions:** This research discovered frequently appearing medical problems across five stages of CKD and further showed that many of the markers reported in previous studies are more applicable to African American patients than Caucasian patients.

Keywords: Machine Learning, Electronic Health Records, Renal Insufficiency, Kidney Failure, Glomerular Filtration Rate

Submitted: August 12, 2017

Revised: 1st, October 13, 2017; 2nd, October 20, 2017

Accepted: October 20, 2017

Corresponding Author

Yong-Mi Kim, PhD

School of Library and Information Studies, University of Oklahoma,
Schusterman Center, 4502 East 41st Street, Tulsa, OK 74135, USA.
Tel: +1-918-660-3364, E-mail: yongmi@ou.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2017 The Korean Society of Medical Informatics

I. Introduction

Currently, about 1 in 4 adults with diabetes has kidney disease in the United States, and diabetes is the leading cause of end-stage kidney diseases (ESRD) [1,2]. Hypertension often coexists with diabetes mellitus and is a contributor to kidney damage in these patients [3]. The damaged kidney cannot properly filter waste and water out of the body, control blood pressure, or make vital hormones such as erythropoietin (EPO), which is critical for the production of red blood cells [4] and 1,25-dihydroxyvitamin D, which is the active form of vitamin D needed for bone health. Because diabetic patients are likely to develop ESRD, it is imperative to discover

which elements of diabetic patients' medical problems lead to ESRD [1] so that those problems can be targeted for prevention.

To discover the major problems that may lead to ESRD, we used the estimated glomerular filtration rate (eGFR) to classify chronic kidney disease (CKD) (Table 1). The eGFR is calculated using various mathematical formulas, the most common one being the Modification of Diet in Renal Disease formula [5]. By categorizing the progression of CKD, one can observe major medical problems appearing in different stages and identify those that lead to ESRD. Most laboratories do not report eGFR values greater than 60 as an independent variable; therefore, we treated them as one group.

Recognizing the importance of preventive measures, scholars have explored markers of medical problems that occur with various categories of eGFR [6,7]. Markers that are commonly measured as predictors of ESRD are urine albumin excretion and urine albumin/creatinine ratio [8]; glycosylated hemoglobin (HbA1c) and serum phosphorus [9]; glucose and high body mass index (BMI) [10]; HbA1c, BMI, systolic and diastolic hypertension, creatinine, proteinuria, obesity, and smoking [11,12]; and hemoglobin [12].

It is clear from a review of previous research that different studies have identified different markers. These differences could be attributed to research design and sampling differences. More specifically, previous studies have been based on the hypothesis driven research method, meaning that scholars have relied on previous studies for their hypothesis development. Additionally, results could have been obtained from subpopulation research samples. CKD varies greatly across subpopulation groups [11], and different subgroups can result in different markers.

Acknowledging these limitations of previous research, this study employed a data-driven approach, which does

Table 1. eGFR categorization

eGFR	Description
≥60	Normal through mildly decreased
30–59	Moderately decreased
15–29	Severely decreased
<15	Kidney failure
<15 (on dialysis)	Kidney failure

GFR source from the chronic kidney disease stages of the National Kidney Foundation (<https://www.kidney.org/atoz/content/gfr>).

eGFR: estimated glomerular filtration rate.

not require a priori knowledge. To discover markers in each stage and predict those markers leading to ESRD, this study categorized kidney disease stages based on eGFR calculated by using the Modification of Diet in Renal Disease formula. Also, African American and Caucasian patients were chosen as subgroups for the focus of this study because previous studies have consistently reported that CKD between these two groups exhibits exemplary health inequity, and they have different markers [8]; however, little systematic research has been done to directly compare those markers.

The purposes of this study, therefore, were (1) to compare how markers differ between African American and Caucasian type 2 diabetes (T2D) patients; (2) to investigate frequently appearing markers in each stage and thus discover the markers leading to ESRD. Note that big data analytics does not require a priori knowledge and variable selections in advance; therefore, a literature review is not included in the paper. However, existing studies are integrated into the explanation and interpretation of the findings.

II. Methods

1. The Data

Electronic Medical Records (EMRs) can be reliably used to identify people with CKD [11]; therefore, this study used EMR data in the Cerner database. The data extraction was based on the International Classification of Diseases, 9th revision (ICD-9, code 250.00–250.93). The data warehouse is completely compliant with the Health Insurance Portability and Accountability Act (HIPAA) and contains comprehensive electronic health records (e.g., incident, lab, procedure, etc.). Because this dataset had already been collected by a third party, the Institutional Review Board (IRB) for the Protection of Human Subjects determined that this study did not meet the criteria to be considered human subject research.

Because the purpose of the study was to investigate CKD-related markers among T2D patients, the data of all the T2D patients in the data warehouse were retrieved based on the ICD-9 (code 250.00–250.93) recorded in 2012. The 2012 data was the most recently completed EMR data when this research data was collected in 2016. The data was filtered to include those who were outpatients, were African American or Caucasian, had an available eGFR value, and were on dialysis.

2. Data Processing and Coding

Figure 1 shows the process employed to extract the data for

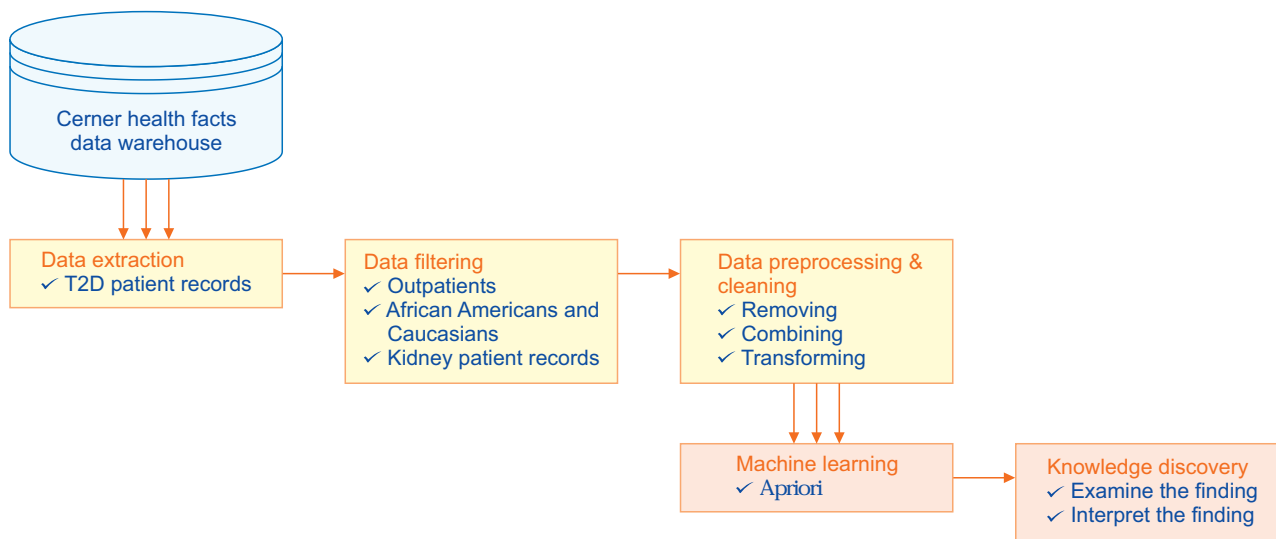


Figure 1. Graphic depiction of the knowledge discovery process.

this research. The total number of retrieved diabetic patient encounters in 2012 was 1,038,499. Because of the wide variety of medical conditions between inpatients and outpatients, we only included outpatient records. This filtering process left 554,732 records. Since the purpose of this study was to draw comparisons between African American and Caucasian patients, other races were excluded, which left 475,656 records. For those two racial groups, there were 354,995 Caucasian patients (74.63%) and 120,661 African American patients (25.37%). The data was then further filtered to include only those patients who had an eGFR value available and those who were on dialysis, which left 4,623 records (Table 2).

3. Analytical Strategy

Researchers choose different machine learning techniques according to their research purposes. For example, a naïve-Bayes classifier is useful when researchers deal with multiple classes to predict the dependent variable, such as detecting cardiovascular disease risk levels [13], and decision trees are useful for splitting and categorizing variables to identify the groups associated with the dependent variable, such as predicting risk factors associated with pressure ulcers [14]. On the other hand, a priori machine learning is suitable for discovering markers because it is designed to discover frequently appearing incidents or sets of incidents. It is an unsupervised machine-learning algorithm that searches all the variables in the database and retrieves only frequently appearing itemsets. As such, it does not require prior knowledge or pre-selected variables. Note that frequently appearing itemsets can be subjective; thus, this technique allows researchers to set the selection criteria using ‘support’, ‘con-

Table 2. Kidney patient distribution by category

eGFR	African American	Caucasian	Total
≥60	634 (23.84)	2,025 (76.16)	2,659
30–59	173 (15.31)	957 (84.69)	1,130
15–29	98 (29.34)	236 (70.66)	334
<15	70 (27.13)	188 (72.87)	258
<15 (on dialysis)	98 (40.50)	144 (59.50)	242
Total	1,073 (23.21)	3,550 (76.79)	4,623

Values are presented as number (%).

eGFR: estimated glomerular filtration rate.

fidence’, and ‘lift’ values and to use them for the validation of findings. Calculations and explanations with examples of those terms are provided in Table 3. To explain the calculations, an example is shown in Table 4, and the calculation method is provided in Table 3.

The dataset used in this study included 3,237 possible medical problems (3,194 lab results and 43 events) from which this machine learning technique can discover rules. The pruning criteria was set using support = 0.00001 and confidence = 1.00. A low support level was set because of the presence of a large set of variables (i.e., 3,237). Because certain sets of medical problems are highly likely to occur in a specific eGFR category, the high confidence level was set. Although it is traditional to report individual values of ‘support’, ‘confidence’, and ‘lift’, if we report all the values of the discovered 1,724 rules, it becomes meaningless. Furthermore, the purpose of this study was to discover markers in

Table 3. Support, confidence, and lift calculation for patient with diabetes

Concept	Formula	Application	Value
Support ^a	$\text{sup}(X) = \frac{(X \cup Y)}{N}$	$\frac{\{\text{high BMI, smoking, high blood pressure}\}}{5}$	$\frac{3}{5} = 0.60$
Confidence ^b	$\text{conf}(X \rightarrow Y) = \frac{(X \cup Y)}{X}$	$\frac{\{\text{high BMI, smoking, high blood pressure}\}}{\{\text{high BMI, smoking}\}}$	$\frac{3}{3} = 1$
Lift ^c	$\text{lift}(X \rightarrow Y) = \frac{\text{sup}(X \cup Y)}{\text{sup}(X) * \text{sup}(Y)}$	$\frac{\{\text{high BMI, smoking, high blood pressure}\}}{\{\text{high BMI, smoking}\} * \{\text{high blood pressure}\}}$	$\frac{0.6}{0.6 * 0.8} = 1.25$

X = {high BMI, smoking}, Y = {high blood pressure}.

^aSupport: the fraction of transactions that contains an itemset. ^bConfidence: how often items in Y appear in hospital visits that contain X. ^cLift: how many more times than expected that XX and YY occurred together if they were statistically independent. A lift value of 1 indicates independence between XX and YY.

Table 4. Example of a priori dataset

Transaction (T) (VisitID)	Itemset (I) (Medical issues)
1	High BMI, smoking, high blood pressure, alcohol use, low sodium
2	High BMI, gastrointestinal, low sodium
3	High BMI, smoking, high blood pressure, GI, low sodium
4	High blood pressure, GI, alcohol use, low sodium, smoking
5	High BMI, smoking, high blood pressure, GI, alcohol use

BMI: body mass index, GI: gastrointestinal.

each stage; therefore, this study aggregated and counted frequently appearing markers in each category of eGFR without reporting the individual values of those criteria. The overall values were the following: support > 0.00001, confidence = 1.00, and lift > 35.

III. Results

Using the criteria of support = 0.00001 and confidence = 1, a total of 1,676 rules for African American patients and 48 rules for Caucasian patients were discovered (Table 5). No rules were discovered for Caucasian patients at eGFR ≤59 mL/min through dialysis, which means that African American patients have CKD-related medical problems at higher rates at that stage. If the criteria were much lower, medical problems could be found for Caucasian patients; however, the probability of those medical problems occurring would be very low; thus the validation of the findings would be very weak. More specific discussion is provided based on the

Table 5. Discovered unique rules for African American and Caucasian patient groups

eGFR	African American	Caucasian	Total
≥60	744	48	792
30–59	155	0	155
15–29	581	0	581
<15	22	0	22
<15 (on dialysis)	174	0	174
Total	1,676	48	1,724

eGFR: estimated glomerular filtration rate.

subgroups of African American and Caucasian patients.

1. Results for the African American Group

Table 5 shows the rules discovered for each stage. A total of 744 rules were discovered for African Americans with eGFR ≥60 mL/min. Since the rules were so numerous, it would not have been informative to report all of the individual rules with values. An efficient way to report the finding of those rules was to summarize the most frequently appearing medical problems among those 744 rules so markers could be identified. Some medical issues appeared only a few times, providing little insight into identifying markers. As such, the medical issues with a frequency of less than 1% were not included. Table 6 presents a summary of the most frequently appearing medical problems appearing at all stages.

1) Medical problems appearing across all levels of eGFR

The medical problems appearing at all levels of eGFR include high glucose, high systolic blood pressure (BP), obesity, alcohol/drug use, low hematocrit, low hemoglobin, low lymphocytes, and coughing. In subjects with eGFR ≥60 mL/

Table 6. Summary of medical problems for African American

eGFR	Medical problems
All levels of eGFR	High glucose High systolic blood pressure Obese Alcohol/drug use Low hematocrit Low hemoglobin Low lymphocytes Cough
All levels of eGFR except for dialysis	Glucose (low at dialysis) GI (very low at dialysis) HbA1c
Medical problems strongly associated with ESRD	High alkaline phosphatase (ALP) High creatinine High phosphorus
Unique to dialysis	High monocyte Low albumin, high anion gap Night sweats High troponin I Low blood urea nitrogen creatinine ratio Genital Low complete blood count crenated red blood cell High differential anisocytosis BNP-B type natriuretic peptide Nausea and vomiting

eGFR: estimated glomerular filtration rate, ESRD: end-stage kidney disease, GI: gastrointestinal, BNP-B: B-type natriuretic peptide.

min, these symptoms could be related to complications of T2D patients. This study found that hypertension, especially high systolic BP, is a stronger predictor for ESRD than diastolic BP, which is consistent with previous studies [11]. This study further found that high diastolic BP is a very important factor for patients who had eGFR <15 mL/min and were undergoing dialysis.

When the kidney does not properly function, it does not produce the erythropoietin (EPO) hormone. EPO signals the bone marrow to make red blood cells [4]. When there is a deficiency of EPO, the body cannot produce red blood cells, causing low hematocrit and low hemoglobin [12]. The findings of this study are consistent with those of previous reports, in that low hematocrit and low hemoglobin are strong predictors of the development of ESRD even in the

early stages of kidney disease.

Low lymphocyte count was a strong predictor of ESRD in all categories. Lymphocytes are a type of white blood cell that fights against infection. As such, a low lymphocyte level along with low hemoglobin and low hematocrit acts as a signal for kidney problems at an early stage for African American patients.

Unlike previous studies that found serum phosphorus, serum creatinine, albumin, sodium, potassium, and chloride as markers [9,15], this study did not find these elements at this stage. This may be attributed to sampling differences. More specifically, this finding is based on African American patients in the United States, while previous studies were based on Asian patients [15] or a single institution [16]. Coughing is a persistently appearing medical problem, and chronic coughing is observed at the dialysis stage. This problem was not reported by previous studies, and further investigation is required.

2) Medical problems appearing across all levels of eGFR except for dialysis

T2D is associated with an increased prevalence of upper and lower gastrointestinal symptoms, and these symptoms appeared to be independently linked to poor glycemic control measured by the HbA1c levels [17]. HbA1c is an early glycation product used diagnostically as a specific marker for diabetic control. Gastrointestinal (GI) symptoms are very common among kidney patients because there is a rise in urea being excreted by the GI tract, increased gastrin levels, and acidosis [18]; however, these issues are less prevalent at the dialysis stage.

3) Medical problems strongly associated with ESRD

Medical problems appearing in the African American group are associated with ESRD, but their appearance is not consistent throughout the stages. These problems include high alkaline phosphatase (ALP), high creatinine, and high phosphorus. ALP is an enzyme found in the bloodstream that is released from the bones and the liver [19]. Elevated ALP levels are associated with kidney disease. In the setting of kidney disease, elevated ALP levels are frequently an indicator of increased bone turnover. Clinicians may use ALP as a risk assessment tool to check patients for the progression of kidney and/or the development of metabolic bone disease [20]. This study confirmed that ALP can be used to predict ESRD for African American patients.

Serum creatinine (SCr) is commonly used among clinicians to determine renal function [21]. It is released by the

muscles and excreted by the kidneys [22]. High SCr appears very frequently at the level of eGFR ≤ 59 mL/min (ranked third), and it was identified as the strongest predictor for ESRD. The findings showed that SCr is consistently higher for African American patients, and it was a strong predictor for the development of ESRD in this dataset.

At eGFR < 15 mL/min, elevated phosphorus levels are observed. High phosphorus occurs because the kidney cannot excrete the mineral from the body. Higher phosphorus for African American patients may be related to cultural eating habits [23]. African Americans consume more fried chicken, sausage, fried fish, and salty snacks than Caucasians [24]. These dietary practices negatively influence health in African American patients undergoing HD. Malnutrition often occurs, and GI symptoms are contributors due to reduced food intake.

4) Medical problems appearing with dialysis

High SCr is a strong predictor for ESRD starting at eGFR ≤ 59 mL/min, and ALP continues to be elevated from the early stage of CKD. Although coughing appears throughout the stages, it worsens when a patient reaches the dialysis stage. High phosphorus started to appear at eGFR < 15 mL/min and slightly slows down at the dialysis stage. This may be due to doctors having advised their patients to reduce their phosphorus intake and/or having prescribed medications to bind phosphorus.

The unique medical problems at this stage are high monocytes, low albumin, high anion gap, night sweats, high troponin I, low blood urea nitrogen (BUN)/SCr, low complete blood count, crenated red blood cells, high differential, anisocytosis, B-type natriuretic peptide, nausea, and vomiting. Albumin is a protein found in the blood and is an indicator of the nutritional status. Malnutrition due to decreased protein intake, inflammation, or comorbidities is common in dialysis patients [25]. Urea nitrogen is formed from the breakdown of protein [26]. Higher than normal levels are usually due to decreased excretion, but they could also be due to excessive catabolism or, sometimes, excessive protein intake [26]. Low BUN/creatinine may be a reflection of dialysis clearing blood of urea nitrogen more effectively than creatinine. A high anion gap appears at the dialysis stage because the kidney cannot excrete acids (sulfate and phosphate) due to the ongoing disease process [27]. Note that the inability to excrete creatinine is the most serious medical problem for dialysis patients.

2. Results for the Caucasian Group

Table 5 shows the numbers of rules discovered for the Caucasian subgroup, which were remarkably different from those of the African American subgroup. Except for subjects with eGFR ≥ 60 mL/min, no rules met the 100% confidence level and the 0.001% support level. As noted, subjects with eGFR ≥ 60 mL/min are usually treated by their primary care doctors rather than nephrologists, and the medical problems appearing at this stage can be related to diabetic aggravators. The notable finding is that smoking is the most significant problem for this group, and alcohol/drug use is ranked 10. Ironically, numerous studies have reported that smoking accelerates the progression of kidney diseases, from early to later stages of diabetes [28], but the findings of this study did not confirm this observation for the Caucasian group, as there were no rules discovered for the categories of eGFR ≤ 59 mL/min through dialysis. Along with smoking, high glucose and obesity failed to emerge as markers for this group.

IV. Discussion

The findings for the African American and Caucasian subgroups were remarkably different. Most notably, the rules discovered using the same criteria differed greatly; the African American group has much higher rates of frequently appearing medical problems across all stages. Because the only rules discovered for the Caucasian group were for the category of eGFR ≥ 60 mL/min, our comparison focused on this stage only. The number of frequently appearing medical problems above 1% for African American patients was 27, while it was 19 for Caucasian patients.

Both groups showed high HbA1c, GI symptoms, high glucose, high triglycerides, obesity, alcohol/drug usage, high and low lymphocytes, low hemoglobin, low hematocrit, cough, and high aspartate aminotransferase. As previous studies have reported, high HbA1c and obesity are strong predictors for ESRD, and these two elements are the most common medical problems for both groups. Hypertension, especially systolic BP, is a risk factor for ESRD, and it is a stronger predictor for African American patients.

A high microalbumin/creatinine ratio was only observed in African American group. This test is commonly performed for diabetic and hypertensive patients to screen for possible kidney damage [29]. It seems that this test would serve as a good predictor for ESRD in African American patients.

Notable differences were also observed between the smoking tendencies of the Caucasian and African American

groups; the latter showed a lower frequency of smoking. Smoking may increase the number of medical problems, but it does not seem to be a strong predictor for ESRD among Caucasian patients; however, further investigation is recommended.

The conclusions of this study may be summarized as follows. (1) Many of the markers reported by previous studies are relevant to African American patients. (2) Systolic BP is a superior predictor for CKD than diastolic BP at the early stages, but diastolic BP was identified as a better predictor at eGFR <15 mL/min. (3) High glucose, obesity, alcohol/drug abuse, and low hematocrit and hemoglobin are strong predictors across all stages for African American patients, unlike Caucasian patients. (4) This study discovered a previously unreported association of certain medical problems among African Americans with ESRD, namely, coughing, night sweats, and high respiratory rate.

Conflict of Interest

No potential conflicts of interest relevant to this article were reported.

Acknowledgments

This project is supported by the College of Arts and Sciences, and the Center for Research Program Development and Enrichment at the University of Oklahoma reflect the views of the Cerner Corporation. The authors greatly appreciate the research assistant, Codi Jones, for editing this manuscript. This work was conducted with data from the Cerner Corporation's Health Facts data warehouse of electronic medical records. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Cerner Corporation.

References

1. Agodoa L, Eggers P. Racial and ethnic disparities in end-stage kidney failure-survival paradoxes in African-Americans. *Semin Dial* 2007;20(6):577-85.
2. National Institute of Diabetes and Digestive and Kidney Diseases. Diabetic kidney disease: What is diabetic kidney disease? [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases; c2017 [cited at 2017 Mar 28]. Available from: [Machine Learning to Compare Frequent Medical Problems](https://www.niddk.nih.gov/health-information/diabetes/over-

</div>
<div data-bbox=)

- view/preventing-problems/diabetic-kidney-disease.
3. National Institute of Diabetes and Digestive and Kidney Diseases. High blood pressure & kidney disease [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases; c2017 [cited at 2017 Mar 28]. Available from: <https://www.niddk.nih.gov/health-information/kidney-disease/chronic-kidney-disease-ckd/high-blood-pressure>.
4. National Institute of Diabetes and Digestive and Kidney Diseases. Anemia and chronic kidney disease [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases; c2017 [cited at 2017 Mar 28]. Available from: <https://www.kidney.org>.
5. Botev R, Mallie JP, Wetzels JF, Couchoud C, Schuck O. The clinician and estimation of glomerular filtration rate by creatinine-based formulas: current limitations and quo vadis. *Clin J Am Soc Nephrol* 2011;6(4):937-50.
6. Dinwiddie LC, Burrows-Hudson S, Peacock EJ. Stage 4 chronic kidney disease: preserving kidney function and preparing patients for stage 5 kidney disease. *Am J Nurs* 2006;106(9):40-51.
7. Lakshman SG, Ravikumar P, Kar G, Das D, Bhattacharjee K, Bhattacharjee P. A comparative study of neurological complications in chronic kidney disease with special reference to its stages and haemodialysis status. *J Clin Diagn Res* 2016;10(12):OC01-OC04.
8. United States Renal Data System. Chapter 2: identification and care of patients with CKD [Internet]. Ann Arbor (MI): United States Renal Data System; c2016 [cited 2017 May 5]. Available from: https://www.usrds.org/2016/view/v1_02.aspx.
9. Ceriello A. The glucose triad and its role in comprehensive glycaemic control: current status, future management. *Int J Clin Pract* 2010;64(12):1705-11.
10. Hintsas S, Dube L, Abay M, Angesom T, Workicho A. Determinants of diabetic nephropathy in Ayder Referral Hospital, Northern Ethiopia: a case-control study. *PLoS One* 2017;12(4):e0173566.
11. de Lusignan S. Informatics as tool for quality improvement: rapid implementation of guidance for the management of chronic kidney disease in England as an exemplar. *Healthc Inform Res* 2013;19(1):9-15.
12. Yamamoto T, Miyazaki M, Nakayama M, Yamada G, Matsushima M, Sato M, et al. Impact of hemoglobin levels on renal and non-renal clinical outcomes differs by chronic kidney disease stages: the Gonryo study. *Clin Exp Nephrol* 2016;20(4):595-602.
13. Miranda E, Irwansyah E, Amelga AY, Maribondang

- MM, Salim M. Detection of cardiovascular disease risk's level for adults using naive Bayes classifier. *Healthc Inform Res* 2016;22(3):196-205.
14. Moon M, Lee SK. Applying of decision tree analysis to risk factors associated with pressure ulcers in long-term care facilities. *Healthc Inform Res* 2017;23(1):43-52.
 15. Tsai CW, Ting IW, Yeh HC, Kuo CC. Longitudinal change in estimated GFR among CKD patients: a 10-year follow-up study of an integrated kidney disease care program in Taiwan. *PLoS One* 2017;12(4):e0173843.
 16. Healthline. ALP (alkaline phosphatase level) test [Internet]. San Francisco (CA): Healthline Media; c2017 [cited at 2017 Mar 28]. Available from: <http://www.healthline.com/health/alp>.
 17. Kim JH, Park HS, Ko SY, Hong SN, Sung IK, Shim CS, et al. Diabetic factors associated with gastrointestinal symptoms in patients with type 2 diabetes. *World J Gastroenterol* 2010;16(14):1782-7.
 18. Idorn T, Knop FK, Jorgensen M, Holst JJ, Hornum M, Feldt-Rasmussen B. Gastrointestinal factors contribute to glucometabolic disturbances in nondiabetic patients with end-stage renal disease. *Kidney Int* 2013;83(5):915-23.
 19. Taliericio JJ, Navaneethan SD. Serum alkaline phosphatase has prognostic importance in chronic kidney disease: increasing levels signal higher risks of ESRD, mortality [Internet]. Cleveland (OH): Cleveland Clinic; c2014 [cited at 2017 Mar 28]. Available from: <https://consultqd.clevelandclinic.org/2014/12/serum-alkaline-phosphatase-has-prognostic-importance-in-chronic-kidney-disease/>.
 20. Chang WX, Xu N, Kumagai T, Shiraishi T, Kikuyama T, Omizo H, et al. The impact of normal range of serum phosphorus on the incidence of end-stage renal disease by a propensity score analysis. *PLoS One* 2016;11(4):e0154469.
 21. DSa J, Shetty S, Bhandary RR, Rao AV. Association between serum cystatin C and creatinine in chronic kidney disease subjects attending a tertiary health care centre. *J Clin Diagn Res* 2017;11(4):BC09-BC12.
 22. National Kidney Foundation. African Americans and kidney disease [Internet]. New York (NY): National Kidney Foundation; c2017 [cited at 2017 Mar 29]. Available from: <https://www.kidney.org/atoz/content/AfricanAmericans-KD>.
 23. Tussing-Humphreys LM, Thomson JL, Onufrak SJ. A church-based pilot study designed to improve dietary quality for rural, lower Mississippi Delta, African American adults. *J Relig Health* 2015;54(2):455-69.
 24. Tucker KL, Maras J, Champagne C, Connell C, Goolsby S, Weber J, et al. A regional food-frequency questionnaire for the US Mississippi Delta. *Public Health Nutr* 2005;8(1):87-96.
 25. Zyga S, Christopoulou G, Malliarou M. Malnutrition-inflammation-atherosclerosis syndrome in patients with end-stage renal disease. *J Ren Care* 2011;37(1):12-5.
 26. MedlinePlus. BUN - blood test [Internet]. Bethesda (MD): US National Library of Medicine; c2017 [cited at 2017 Mar 28]. Available from: <https://medlineplus.gov/ency/article/003474.htm>.
 27. Kaslow J. Anion gap [Internet]. Santa Ana (CA): drkaslow.com; c2017 [cited at 2017 Mar 28]. Available from: http://www.drkaslow.com/html/anion_gap.html.
 28. Orth SR. Effects of smoking on systemic and intrarenal hemodynamics: influence on renal function. *J Am Soc Nephrol* 2004;15 Suppl 1:S58-63.
 29. American Association for Clinical Chemistry. Urine albumin and albumin/creatinine ratio [Internet]. Washington (DC): American Association for Clinical Chemistry; c2017 [cited at 2017 May 23]. Available from: <https://labtestsonline.org/understanding/analytes/microalbumin/tab/test/>.