




ORIGINAL ARTICLE

Subtyping hospitalized patients with hypokalemia by machine learning consensus clustering and associated mortality risks

Charat Thongprayoon ¹, Michael A. Mao², Andrea G. Kattah¹, Mira T. Keddis ³, Pattharawin Pattharanitima⁴, Stephen B. Erickson¹, John J. Dillon¹, Vesna D. Garovic¹ and Wisit Cheungpasitporn ¹

¹Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Rochester, MN, USA,

²Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Jacksonville, FL, USA,

³Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Phoenix, AZ, USA and

⁴Department of Internal Medicine, Faculty of Medicine, Thammasat University, Pathum Thani, Thailand

Correspondence to: Wisit Cheungpasitporn; E-mail: wcheungpasitporn@gmail.com

ABSTRACT

Background. Hospitalized patients with hypokalemia are heterogeneous and cluster analysis, an unsupervised machine learning methodology, may discover more precise and specific homogeneous groups within this population of interest. Our study aimed to cluster patients with hypokalemia at hospital admission using an unsupervised machine learning approach and assess the mortality risk among these distinct clusters.

Methods. We performed consensus clustering analysis based on demographic information, principal diagnoses, comorbidities and laboratory data among 4763 hospitalized adult patients with admission serum potassium ≤ 3.5 mEq/L. We calculated the standardized mean difference of each variable and used the cutoff of ± 0.3 to identify each cluster's key features. We assessed the association of the hypokalemia cluster with hospital and 1-year mortality.

Results. Consensus cluster analysis identified three distinct clusters that best represented patients' baseline characteristics. Cluster 1 had 1150 (32%) patients, cluster 2 had 1344 (28%) patients and cluster 3 had 1909 (40%) patients. Based on the standardized difference, patients in cluster 1 were younger, had less comorbidity burden but higher estimated glomerular filtration rate (eGFR) and higher hemoglobin; patients in cluster 2 were older, more likely to be admitted for cardiovascular disease and had higher serum sodium and chloride levels but lower eGFR, serum bicarbonate, strong ion difference (SID) and hemoglobin, while patients in cluster 3 were older, had a greater comorbidity burden, higher serum bicarbonate and SID but lower serum sodium, chloride and eGFR. Compared with cluster 1, cluster 2 had both higher hospital and 1-year mortality, whereas cluster 3 had higher 1-year mortality but comparable hospital mortality.

Conclusion. Our study demonstrated the use of consensus clustering analysis in the heterogeneous cohort of hospitalized hypokalemic patients to characterize their patterns of baseline clinical and laboratory data into three clinically distinct clusters with different mortality risks.

Received: 10.5.2021; Editorial decision: 20.9.2021

© The Author(s) 2021. Published by Oxford University Press on behalf of the ERA. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Keywords: artificial intelligence, clustering, electrolytes, hypokalemia, machine learning, potassium

INTRODUCTION

Hypokalemia, defined as a serum potassium level ≤ 3.5 mEq/L, is a common clinical problem detected in 2–3% of outpatient encounters [1–3] and up to 21% of hospitalized patients [3–5]. Hypokalemia promotes cardiac arrhythmias by reducing repolarization reserve and increasing intracellular Ca^{2+} in cardiac myocytes [6]. This effect paradoxically increases the excitability of cardiac myocytes, predisposing to ventricular arrhythmias [6]. Even mild asymptomatic hypokalemia (defined as a serum potassium of 2.5–3.4 mEq/L) has been associated with adverse clinical outcomes, including increased cardiovascular events, hospital mortality and 1-year mortality [1, 4, 7, 8].

Studies have demonstrated that certain subgroups of hospitalized hypokalemic patients carry an even greater risk of mortality regardless of cause, such as patients with cardiovascular diseases (CVDs) or chronic kidney diseases (CKDs) [4, 9]. However, not all patients with CVD or CKD have similar characteristics, and there are many important patient-associated factors and coexisting illnesses that may affect patient survival [10]. Patients with CVD or CKD are a heterogeneous group with variable risk associations with cardiovascular events and mortality. Studies have demonstrated that machine learning (ML) cluster analysis may identify meaningful disease subtypes and/or groups of related phenotypic variables, even in a highly selected group of patients [11, 12]. Given that hospitalized patients with hypokalemia are heterogeneous [4, 7], cluster analysis, an unsupervised ML methodology, may discover more precise and specific homogeneous groups within this population of interest [13, 14]. The clustering approach can identify subgroups/clusters of patients with homogeneous values incorporating all variables, not just the target variable [15–18]. Thus the utilization of clustering analysis on a large hospitalized patient cohort with hypokalemia may enable us to identify clusters of patients sharing different phenotypic and clinicopathological features, which may be associated with different clinical outcomes including mortality.

In this study we aimed to explore the clustering of patients with hypokalemia on hospital admission and assess the mortality risk among these distinct clusters.

MATERIALS AND METHODS

Patient population

The Mayo Clinic Institutional Review Board approved this study. We used our hospital's database to identify all adult patients (≥ 18 years) admitted to the Mayo Clinic Hospital from 1 January 2011 to 31 December 2013. If patients had more than one hospital admission during the study period, we analyzed only the first admission. We included patients who had hypokalemia at hospital admission. We defined admission hypokalemia as the first serum potassium measured within 24 hours of hospital admission ≤ 3.5 mEq/L. We excluded patients who did not have a serum potassium measurement within 24 hours of hospital admission and patients who did not provide research authorization.

Data collection

We abstracted pertinent clinical characteristics and laboratory data from our hospital's electronic database, using a previously validated method [4, 7, 19, 20]. In brief, the Mayo

Clinic Life Science System database contains demographic characteristics, hospital admissions information, diagnosis and procedure codes, laboratory test results and flow sheet data of inpatients and outpatients. Principal diagnoses were grouped based on International Classification of Diseases, 9th Revision (ICD-9) codes at admission (Supplementary data, Table S1). The Charlson comorbidity index was calculated to determine the comorbidity burden at the time of admission. The comorbid conditions were extracted from clinical notes contained within the electronic medical record using an automated electronic search algorithm [21]. Glomerular filtration rate (GFR) was estimated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation based on age, sex, race and admission serum creatinine [22]. The anion gap (mEq/L) was calculated by serum sodium-(serum chloride + serum bicarbonate) [23]. The strong ion difference (SID; mEq/L) was calculated by (serum sodium + serum potassium)-serum chloride [23].

As our goal was to group hospitalized hypokalemic patients into clusters based on their hospital admission's demographic information, principal diagnoses, comorbidities and laboratory data, we only utilized data that were available within 24 hours of hospital admission for clustering analysis. We selected the first laboratory value within this 24-hour time frame if there were multiple measurements. We excluded variables with $>20\%$ missing data. If variables had missing data accounting for $<20\%$, we imputed missing data through multiple imputations using random forest before their inclusion in clustering analysis.

Clustering analysis

We applied an unsupervised ML approach to develop clinical phenotypes of hospitalized patients with hypokalemia by conducting unsupervised consensus clustering [13]. We used a pre-specified subsampling parameter of 80% with 100 iterations and assigned the number of potential clusters (k) to range from 2 to 10 in order to avoid producing an excessive number of clusters that would not be clinically useful. The optimal number of clusters was determined by examining the consensus matrix (CM) heat map, cumulative distribution function (CDF), cluster-consensus plots in the within-cluster consensus scores and the proportion of ambiguously clustered pairs (PAC) [24, 25]. The within-cluster consensus score, ranging between 0 and 1, is defined as the average consensus value for all pairs of individuals belonging to the same cluster [25]. A value closer to 1 indicates better cluster stability [25]. PAC, ranging between 0 and 1, is calculated as the proportion of all sample pairs with consensus values falling within the predetermined boundaries [24]. A value closer to 0 indicates better cluster stability [24]. We calculated the PAC using two criteria: the strict criteria consisting of a predetermined boundary of (0, 1), where a pair of individuals who had a consensus value >0 or <1 was considered ambiguously clustered, and the relaxed criteria consisting of a predetermined boundary of (0.1, 0.9), where a pair of individuals who had a consensus value >0.1 or <0.9 was considered ambiguously clustered [24]. This study's detailed consensus cluster algorithms are provided in the online supplementary data.

Statistical analysis

After cluster identification, we performed analyses to explore differences among the clusters. We tested differences in

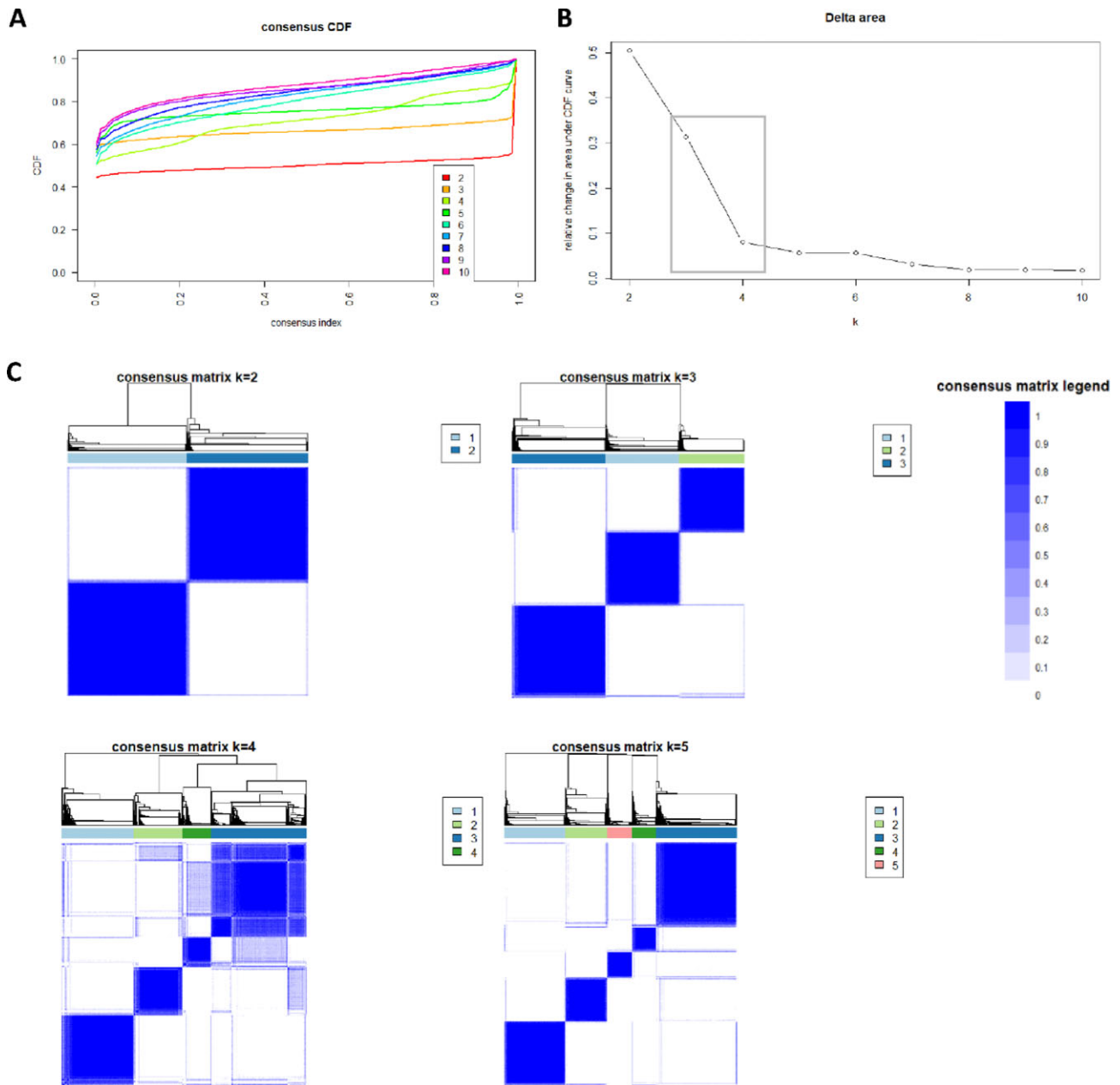


FIGURE 1: (A) CDF plot displaying consensus distributions for each number of clusters (k). (B) Delta area plot reflecting the relative changes in the area under the CDF curve. (C) Consensus matrix heat map depicting consensus values on a white to blue color scale of each cluster. The square demonstrates largest changes in area occurred between $k = 3$ and $k = 4$.

baseline characteristics among the identified clusters using the analysis of variance (ANOVA) test for continuous variables and chi-squared test for categorical variables. To examine cluster profiles, we calculated and graphically displayed the standardized mean differences of baseline characteristics between each cluster and the overall population. We considered variables with an absolute standardized mean difference >0.3 as key features for each cluster. We hypothesized that valid, clinically distinct clusters would have measurable differences in outcomes. Hence we compared hospital mortality and 1-year mortality as outcomes of interest across the clusters. We also examined the association of clusters with hospital mortality using logistic regression and 1-year mortality using Cox proportional hazards regression. We used cluster 1 as the reference group for

outcome comparison, as this cluster was associated with the lowest mortality. We did not adjust for between-group differences in baseline characteristics because these variables were used to develop the clusters through unsupervised ML. We performed all analyses using R version 4.0.3 (RStudio, Boston, MA; <http://www.rstudio.com/>), with the packages of ConsensusClusterPlus (version 1.46.0) [25] for consensus clustering analysis and the missForest package for missing data imputation [26].

RESULTS

There were 147 358 hospital admissions during the study period. Of these, 31 110 were excluded due to no admission serum potassium and 39 552 due to readmission. As such, there were

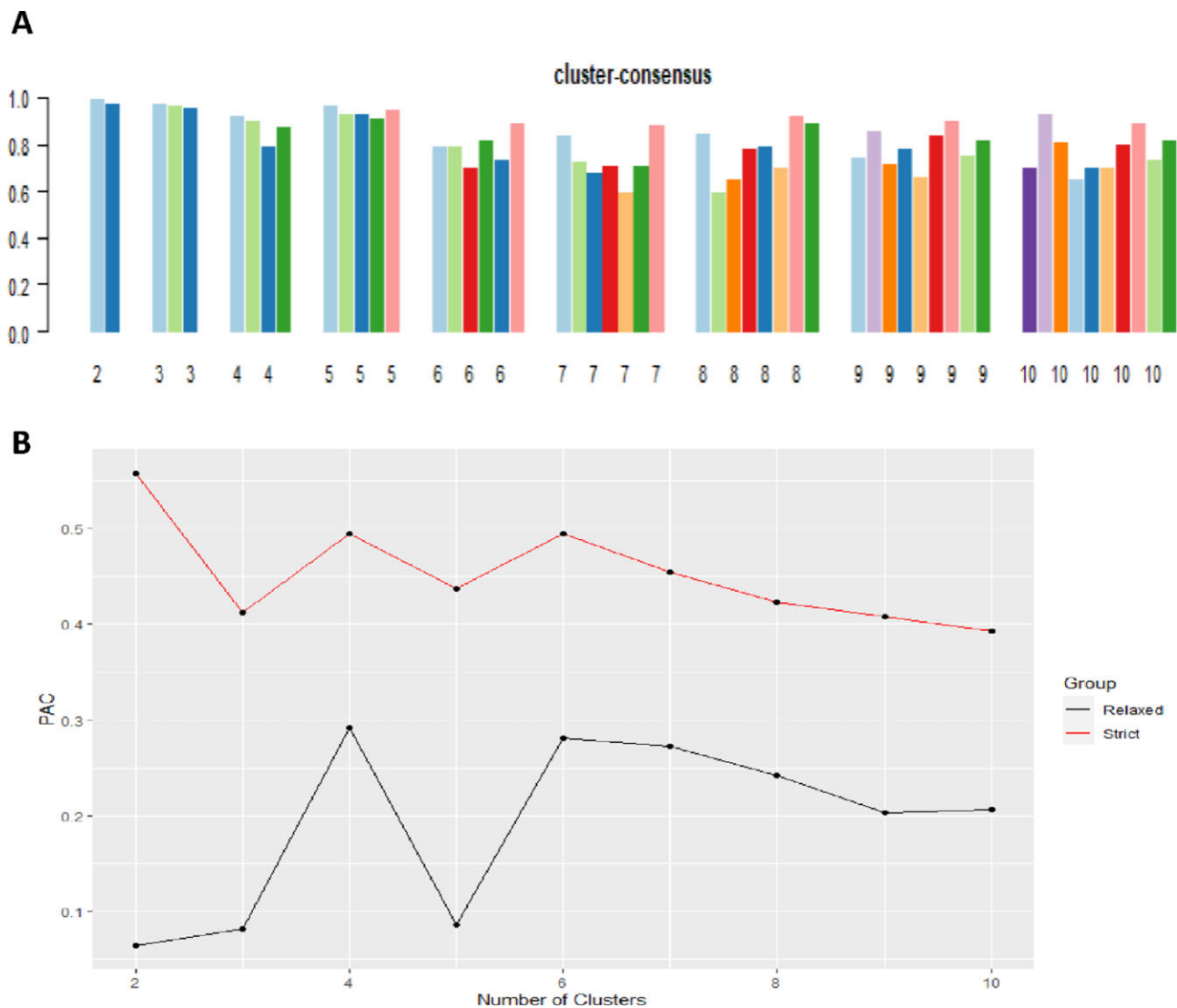


FIGURE 2: (A) The bar plot represents the mean consensus score for different numbers of clusters (k ranges from 2 to 10). (B) The PAC values using the strict criteria (red line) with the predetermined boundary of (0, 1) and the PAC values using the relaxed criteria (black line) with the predetermined boundary of (0.1, 0.9) as the definition for ambiguously clustered pairs.

76696 unique hospitalized patients with available admission serum potassium. We included a total of 4763 hospitalized patients with admission hypokalemia. The mean age was 60 ± 19 years and 39% were male. The mean estimated GFR (eGFR) was 80 ± 24 mL/min/1.73 m² and 19% had CKD. The mean admission serum potassium was 3.2 ± 0.3 mEq/L.

The CDF plot displayed consensus distributions for each potassium level (Figure 1A). The delta area plot shows the relative change in the area under the CDF curve (Figure 1B). The largest changes in area occurred between $k = 3$ and $k = 4$, at which point the relative increase in area became noticeably smaller. As shown in the CM heatmap (Figure 1C; supplementary data, Figures S1–S9), the ML algorithm identified cluster 2 and cluster 3 with clear boundaries, indicating good cluster stability over repeated iterations. The mean cluster-consensus score was comparable between the scenario 2 and 3 clusters (Figure 2A). Favorable low PACs by both strict and relaxed criteria were demonstrated for three clusters (Figure 2B). Thus, using base-

line variables at hospital admission, the consensus clustering analysis identified three clusters that best represented the data pattern of our hospitalized hypokalemic patients.

Cluster 1 had 1150 (32%) patients, cluster 2 had 1344 (28%) patients and cluster 3 had 1909 (40%) patients. Table 1 shows the baseline characteristics of the three identified clusters. The distribution of all baseline variables was significantly different between the three clusters. Of note, the mean admission serum potassium was numerically comparable across the three clusters [3.2 [standard deviation (SD) 0.2], 3.1 [SD 0.3] and 3.2 [SD 0.3] mEq/L for cluster 1, 2 and 3 respectively], although there was a statistically significant difference ($P < 0.001$). Figure 3 shows the plot of the standardized mean difference to visualize the key features for each cluster. Cluster 1 patients were younger, with a lower comorbidity burden and a higher eGFR and hemoglobin. Cluster 2 patients were older, more likely to be admitted for cardiovascular diagnosis and had higher serum sodium and chloride levels but lower eGFR, serum

Table 1. Clinical characteristics at hospital admission according to clusters in hospitalized patients with hypokalemia

Patient characteristics	Overall (n = 4763)	Cluster 1 (n = 1510)	Cluster 2 (n = 1344)	Cluster 3 (n = 1909)	P-value
Age (years), mean ± SD	59.9 ± 18.7	40.9 ± 13.4	66.8 ± 14.0	70.0 ± 13.1	<0.001
Male sex, n (%)	1880 (39)	659 (44)	498 (37)	723 (38)	<0.001
Race, n (%)					<0.001
White	4308 (90)	1262 (84)	1244 (93)	1802 (94)	
Black	142 (3)	94 (6)	23 (2)	25 (1)	
Others	313 (7)	154 (10)	77 (6)	82 (4)	
BMI (kg/m ²), mean ± SD	28.5 ± 7.3	27.5 ± 7.0	28.5 ± 6.5	29.3 ± 7.9	<0.001
Principal diagnosis, n (%)					<0.001
Cardiovascular	993 (21)	137 (9)	484 (36)	372 (19)	
Endocrine/metabolic	227 (5)	58 (4)	39 (3)	130 (7)	
Gastrointestinal	628 (13)	204 (14)	134 (10)	290 (15)	
Kidney and urinary	142 (3)	37 (2)	33 (2)	72 (4)	
Hematology/oncology	413 (9)	73 (5)	150 (11)	190 (10)	
Infectious disease	337 (7)	114 (8)	96 (7)	127 (7)	
Respiratory	242 (5)	65 (4)	39 (3)	138 (7)	
Injury/poisoning	852 (18)	428 (28)	191 (14)	233 (12)	
Other	929 (20)	394 (26)	178 (13)	357 (19)	
Charlson comorbidity score, mean ± SD	1.7 ± 2.4	0.6 ± 1.2	1.9 ± 2.4	2.5 ± 2.7	<0.001
Comorbidities, n (%)					
Coronary artery disease	277 (6)	26 (2)	94 (7)	157 (8)	<0.001
Congestive heart failure	358 (8)	29 (2)	127 (9)	202 (11)	<0.001
Peripheral vascular disease	133 (3)	2 (0.1)	46 (3)	85 (4)	<0.001
Dementia	68 (1)	1 (0.1)	22 (2)	45 (2)	<0.001
Stroke	378 (8)	16 (1)	117 (9)	245 (13)	<0.001
COPD	399 (8)	31 (2)	101 (8)	267 (14)	<0.001
Diabetes mellitus	823 (17)	86 (6)	281 (21)	456 (24)	<0.001
Cirrhosis	150 (3)	32 (2)	51 (4)	67 (4)	0.02
Chronic kidney disease	894 (19)	7 (0.5)	387 (29)	500 (26)	<0.001
End-stage kidney disease	120 (3)	7 (0.5)	51 (4)	62 (3)	<0.001
Laboratory tests, mean ± SD					
eGFR (mL/min/1.73 m ²)	80 ± 24	102 ± 16	69 ± 19	69 ± 19	<0.001
Sodium (mEq/L)	138 ± 5	138 ± 4	140 ± 4	136 ± 6	<0.001
Potassium (mEq/L)	3.2 ± 0.3	3.2 ± 0.2	3.1 ± 0.3	3.2 ± 0.3	<0.001
Chloride (mEq/L)	102 ± 7	103 ± 5	109 ± 5	97 ± 6	<0.001
Bicarbonate (mEq/L)	25 ± 5	24 ± 4	22 ± 4	28 ± 4	<0.001
Anion gap (mEq/L)	10 ± 4	11 ± 4	9 ± 4	11 ± 4	<0.001
Strong ion difference (mEq/L)	38.4 ± 5.0	37.6 ± 3.5	34.6 ± 4.4	41.8 ± 4.2	<0.001
Hemoglobin (g/dL)	12.0 ± 2.2	12.8 ± 2.1	10.7 ± 2.0	12.3 ± 2.1	<0.001

COPD, chronic obstructive pulmonary disease; BMI, body mass index.

bicarbonate, SID and hemoglobin. Cluster 3 patients were older, had a higher comorbidity burden, serum bicarbonate and SID but lower serum sodium, chloride and eGFR.

Hospital mortality was 1.9% for cluster 1, 4.2% for cluster 2 and 2.3% for cluster 3 ($P < 0.001$) (Figure 4A). Cluster 2 [odds ratio [OR] 2.34 [95% confidence interval (CI) 1.48–3.71]], but not cluster 3, was significantly associated with higher hospital mortality compared with cluster 1 (Table 2). One-year mortality was 8.0% for cluster 1, 21.4% for cluster 2 and 21.9% for cluster 3 ($P < 0.001$) (Figure 4B). Both cluster 2 [OR 2.87 (95% CI 2.24–3.67)] and cluster 3 [OR 2.94 (95% CI 2.33–3.71)] were significantly associated with increased 1-year mortality compared with cluster 1 (Table 2).

DISCUSSION

In this study we divided hospitalized patients with hypokalemia into three distinct subgroups by applying unsupervised consensus clustering analysis. This resulted in identification of different phenotypic features and prognostic differentiation, including hospital mortality and 1-year mortality. Hospital mortality and 1-year mortality varied among the clusters: 1.9% and 8.0%

in cluster 1, 4.2% and 21.4% in cluster 2 and 2.3% and 21.9% in cluster 3, respectively.

The practice of medicine is advancing to more personalized care. ML methods, such as cluster analysis, offer the ability to more efficiently analyze large volumes of data to identify and classify groups of patients based on phenotypic features, thus allowing the development of targeted interventions to improve patient outcomes. In this study we successfully demonstrated the ability to identify clinically important groupings by unsupervised ML consensus clustering, which may have potential implications for the management of hospitalized patients with hypokalemia.

Hypokalemic patients in cluster 2 had the highest in-hospital mortality among the three groups. Furthermore, patients in this group also had reduced 1-year survival when compared with cluster 1. Characteristics of hypokalemic patients in cluster 2 included older age, principal diagnosis of CVD, anemia, reduced GFR, hyperchloremic metabolic acidosis and hypernatremia (Figure 3). Previous studies with traditional statistics have demonstrated that hypokalemia is associated with increased risks of in-hospital mortality, ventricular arrhythmias and

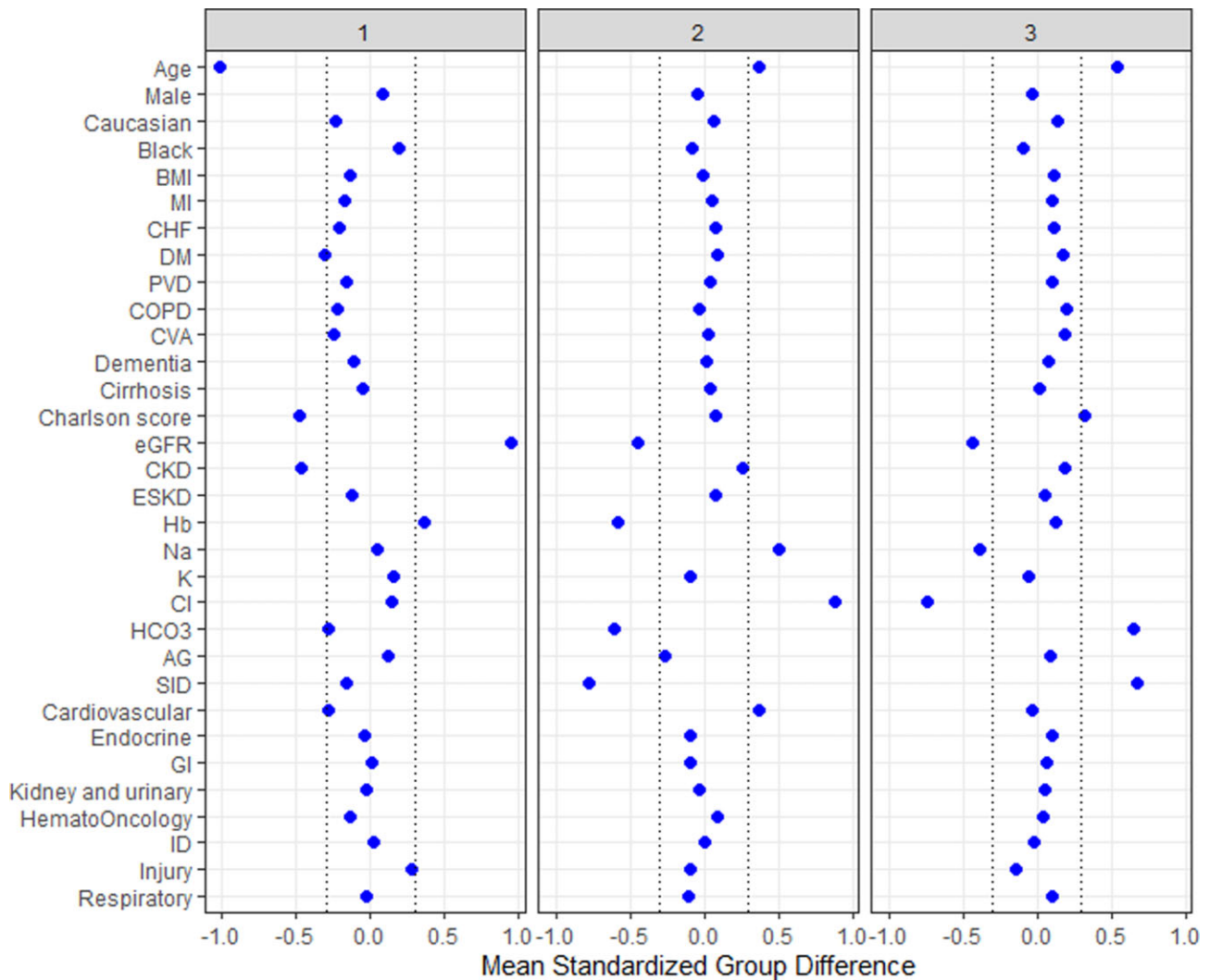


FIGURE 3: The standardized differences across three clusters for each of the baseline parameters. The x axis is the standardized differences value and the y axis shows baseline parameters. The dashed vertical lines represent the standardized differences cutoffs of <-0.3 and >0.3 . AG, anion gap; BMI, body mass index; Cl, chloride; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; DM, diabetes mellitus; ESKD, end-stage kidney disease; GI, gastrointestinal; Hb, hemoglobin; ID, infectious disease; HCO_3^- , bicarbonate; K, potassium; MI, myocardial infarction; Na, sodium; PVD, peripheral vascular disease.

cardiac arrest, especially among CVD and CKD patient populations [4, 7, 27–32]. In addition to CKD and CVD, additional phenotypic features of patients in cluster 2 included anemia, hyperchloremic metabolic acidosis and hypernatremia. Given that patients with these phenotypes carried higher in-patient and 1-year mortality, future studies are needed to identify strategies to identify those at risk to improve outcomes among this patient population.

Cluster 3 consists of older patients with a high comorbidities burden and an eGFR that falls between cluster 1 and cluster 2. These patients also had high SID (increased HCO_3^- or decreased H^+), which usually occurs in the settings of serum chloride loss or serum sodium gain. Since serum sodium in this cluster is lower than in other clusters, these hypokalemic patients in cluster 3 likely had phenotypes of metabolic alkalosis with chloride depletion. This phenotype of metabolic alkalosis with chloride depletion was not identified in the other clusters. Clinically this is often observed in the setting of vomiting/nasogastric suction, bulimia, chloruretic effect of diuretics

and post-hypercapnic state [33, 34]. Despite not reaching statistical significance, patients in cluster 3 had a higher hospital mortality rate compared with patients in cluster 1 (which is comprised of younger adult patients with the lowest comorbidities burden and highest eGFR among the three clusters). These patients in cluster 3 had reduced 1-year survival compared with patients in cluster 1, but comparable to that of patients in cluster 2. Thus hypokalemic patients with phenotypic features of cluster 3 may require close post-hospitalization follow-up to reduce 1-year mortality risk.

There were some limitations to our current study. First, although we used a large dataset, the data from our study are from a single center. Hence the clustering may be unique to our patient population (which is predominantly non-Hispanic white). Second, we performed ML clustering at the time of hospital admission. The workup for hypokalemic patients in the outpatient setting or throughout their hospitalization may include studies that evaluate for potassium transcellular shifts versus urinary loss versus gastrointestinal loss. This

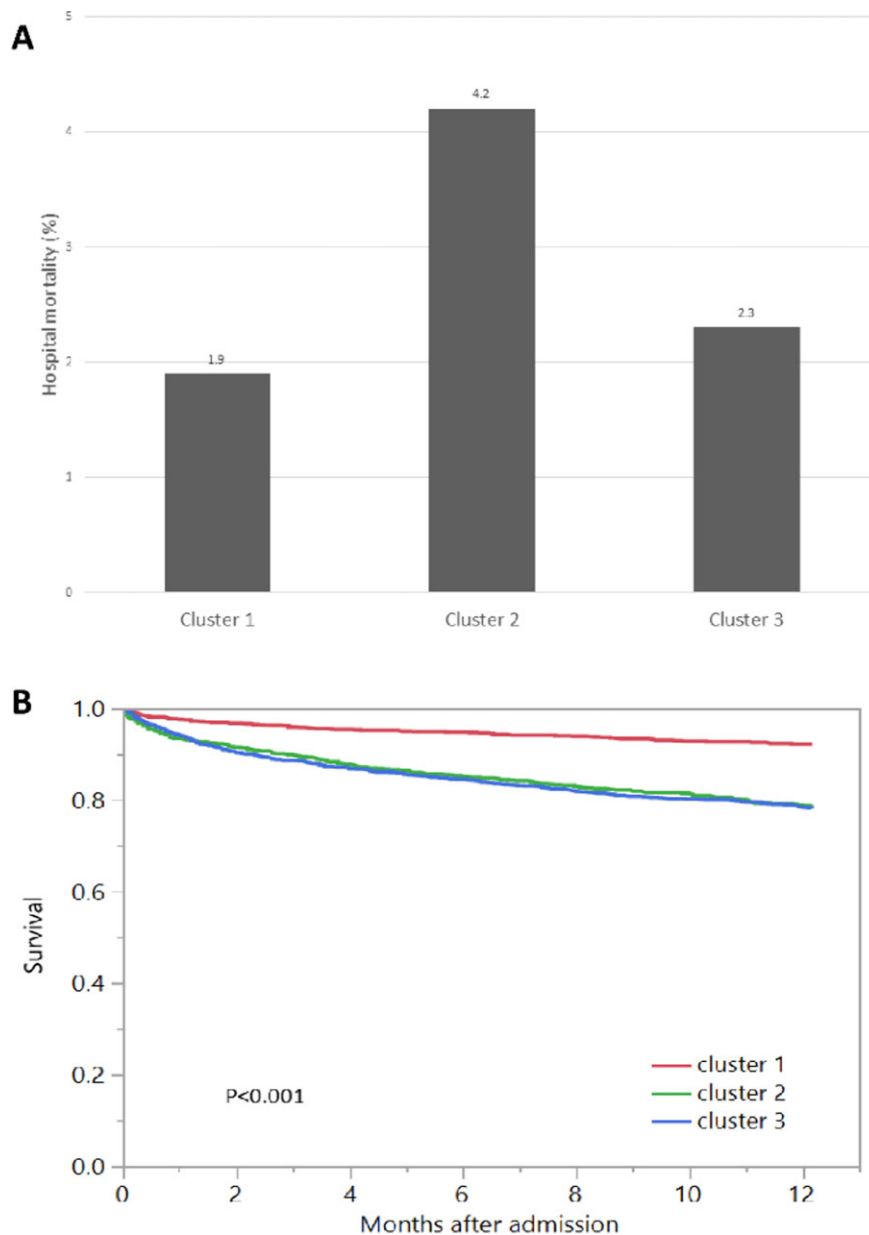


FIGURE 4: (A) Hospital mortality among different clusters of admission hypokalemia. (B) One-year mortality among different clusters of admission hypokalemia.

Table 2. Mortality outcomes according to clusters

Cluster	Hospital mortality, %	OR (95% CI)	1-year mortality, %	HR (95% CI)
1	1.9	1 (ref)	8.0	1 (ref)
2	4.2	2.34 (1.48–3.71)	21.4	2.87 (2.24–3.67)
3	2.3	1.25 (0.77–2.02)	21.9	2.94 (2.33–3.71)

HR, hazard ratio.

testing includes, but is not limited to, plasma aldosterone and renin levels, serum magnesium and calcium levels, spot urine samples for the ratio of urine potassium to urine creatinine or urinary chloride and 24-hour urine studies [35, 36]. Given that these laboratory investigations were not commonly performed on admission (even serum magnesium levels were missing

in >20% of patients), these data were limited and thus were not included in our ML clustering algorithm. However, our clustering approach at the time of hospital admission may provide potential implications for the early management of hospitalized patients with distinct phenotypes of hypokalemia. Lastly, data on electrocardiograms and medications that are

common causes of hypokalemia, such as thiazide and loop diuretics, were limited in our database. Thus future studies are needed to assess whether these variables could have improved the discriminatory ability of the groupings we identified.

In summary, we present the first clustering analysis of hospitalized patients with hypokalemia performed to date. The findings of our study allow selection of phenotypic and clinicopathological features in three unique clusters that may be of consequence for the development of hypokalemia and its associated survival outcomes. Cluster 2 identified hypokalemic patients with a phenotype of older age, principal diagnosis of CVD, anemia, reduced GFR, hyperchloremic metabolic acidosis and hypernatremia that had increased in-hospital and 1-year mortality. Patients in cluster 3 had higher 1-year mortality but comparable hospital mortality compared with cluster 1. These findings may potentially help classify hospitalized patients with hypokalemia on admission, thus translating to an improved personalized precision medicine approach.

SUPPLEMENTARY DATA

Supplementary data are available at [ckj](#) online.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

AUTHORS' CONTRIBUTIONS

All authors have read and agreed to the published version of the manuscript.

CONFLICT OF INTEREST STATEMENT

We do not have any financial or nonfinancial potential conflicts of interest.

REFERENCES

- Kovesdy CP, Matsushita K, Sang Y et al. Serum potassium and adverse outcomes across the range of kidney function: a CKD prognosis consortium meta-analysis. *Eur Heart J* 2018; 39: 1535–1542
- Grams ME, Hoenig MP, Hoom EJ. Evaluation of hypokalemia. *JAMA* 2021; 325: 1216–1217
- Viera AJ, Wouk N. Potassium disorders: hypokalemia and hyperkalemia. *Am Fam Physician* 2015; 92: 487–495
- Cheungpasitporn W, Thongprayoon C, Kittanamongkolchai W et al. Impact of admission serum potassium on mortality in patients with chronic kidney disease and cardiovascular disease. *QJM* 2017; 110: 713–719
- Clase CM, Carrero JJ, Ellison DH et al. Potassium homeostasis and management of dyskalemia in kidney diseases: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int* 2020; 97: 42–61
- Weiss JN, Qu Z, Shivkumar K. Electrophysiology of hypokalemia and hyperkalemia. *Circ Arrhythm Electrophysiol* 2017; 10: e004667
- Thongprayoon C, Cheungpasitporn W, Hansrivijit P et al. Admission serum potassium levels in hospitalized patients and one-year mortality. *Medicines (Basel)* 2019; 7: 2
- Luo J, Brunelli SM, Jensen DE et al. Association between serum potassium and outcomes in patients with reduced kidney function. *Clin J Am Soc Nephrol* 2016; 11: 90–100
- Basnet S, Dhital R, Tharu B et al. Influence of abnormal potassium levels on mortality among hospitalized heart failure patients in the US: data from national inpatient sample. *J Community Hosp Intern Med Perspect* 2019; 9: 103–107
- Sundararajan V, Henderson T, Perry C et al. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004; 57: 1288–1294
- Cho MH, Washko GR, Hoffmann TJ et al. Cluster analysis in severe emphysema subjects using phenotype and genotype data: an exploratory investigation. *Respir Res* 2010; 11: 30
- Zheng Z, Waikar SS, Schmidt IM et al. Subtyping CKD patients by consensus clustering: the Chronic Renal Insufficiency Cohort (CRIC) study. *J Am Soc Nephrol* 2021; 32: 639–653
- Monti S, Tamayo P, Mesirov J et al. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learn* 2003; 52: 91–118
- Thongprayoon C, Nissaisorakarn V, Pattharanitima P et al. Subtyping hyperchloremia among hospitalized patients by machine learning consensus clustering. *Medicina (Kaunas)* 2021; 57: 903
- Ženko B, Džeroski S, Struyf J. Learning predictive clustering rules. In: Bonchi F, Boulicaut JF (eds), *Knowledge Discovery in Inductive Databases*. KDID 2005. Lecture Notes in Computer Science, vol 3933. Berlin: Springer, 2005: 234–250
- Wang M, Abrams ZB, Kornblau SM et al. Thresher: determining the number of clusters while removing outliers. *BMC Bioinformatics* 2018; 19: 9
- Thongprayoon C, Hansrivijit P, Mao MA et al. Machine learning consensus clustering of hospitalized patients with admission hyponatremia. *Diseases* 2021; 9: 54
- Thongprayoon C, Kattah AG, Mao MA et al. Distinct phenotypes of hospitalized patients with hyperkalemia by machine learning consensus clustering and associated mortality risks. *QJM* 2021; doi: 10.1093/qjmed/hcab194
- Thongprayoon C, Cheungpasitporn W, Thirunavukkarasu S et al. Serum potassium levels at hospital discharge and one-year mortality among hospitalized patients. *Medicina (Kaunas)* 2020; 56: 236
- Thongprayoon C, Cheungpasitporn W, Chewcharat A et al. Risk of respiratory failure among hospitalized patients with various admission serum potassium levels. *Hosp Pract* 2020; 48: 75–79
- Singh B, Singh A, Ahmed A et al. Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from electronic medical records. *Mayo Clin Proc* 2012; 87: 817–824
- Matsushita K, Mahmoodi BK, Woodward M et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. *JAMA* 2012; 307: 1941–1951
- Kellum JA, Kramer DJ, Pinsky MR. Strong ion gap: a methodology for exploring unexplained anions. *J Crit Care* 1995; 10: 51–55
- Şenbabaoğlu Y, Michailidis G, Li JZ. Critical limitations of consensus clustering in class discovery. *Sci Rep* 2014; 4: 6207
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; 26: 1572–1573

26. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012; 28: 112–118
27. Goyal A, Spertus JA, Gosch K et al. Serum potassium levels and mortality in acute myocardial infarction. *JAMA* 2012; 307: 157–164
28. Krogager ML, Eggers-Kaas L, Aasbjerg K et al. Short-term mortality risk of serum potassium levels in acute heart failure following myocardial infarction. *Eur Heart J Cardiovasc Pharmacother* 2015; 1: 245–251
29. Joy PS, Kumar G, Olshansky B. Syncope: outcomes and conditions associated with hospitalization. *Am J Med* 2017; 130: 699–706.e6
30. Lu YY, Cheng CC, Chen YC et al. Electrolyte disturbances differentially regulate sinoatrial node and pulmonary vein electrical activity: a contribution to hypokalemia- or hyponatremia-induced atrial fibrillation. *Heart Rhythm* 2016; 13: 781–788
31. Collins AJ, Pitt B, Reaven N et al. Association of serum potassium with all-cause mortality in patients with and without heart failure, chronic kidney disease, and/or diabetes. *Am J Nephrol* 2017; 46: 213–221
32. Zhang Y, Chen P, Chen J et al. Association of low serum potassium levels and risk for all-cause mortality in patients with chronic kidney disease: a systematic review and meta-analysis. *Ther Apher Dial* 2019; 23: 22–31
33. Luke RG, Galla JH. It is chloride depletion alkalosis, not contraction alkalosis. *J Am Soc Nephrol* 2012; 23: 204–207
34. Hodgkin JE, Soeprono FF, Chan DM. Incidence of metabolic alkalemia in hospitalized patients. *Crit Care Med* 1980; 8: 725–728
35. Funder JW, Carey RM, Mantero F et al. The management of primary aldosteronism: case detection, diagnosis, and treatment: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2016; 101: 1889–1916
36. Palmer BF, Clegg DJ. The use of selected urine chemistries in the diagnosis of kidney disorders. *Clin J Am Soc Nephrol* 2019; 14: 306–316