

Hierarchical Clustering Analysis for Predicting 1-Year Mortality After Starting Hemodialysis



Yohei Komaru^{1,2,4}, Teruhiko Yoshida^{1,2,4}, Yoshifumi Hamasaki², Masaomi Nangaku^{1,2} and Kent Doi³

¹Division of Nephrology and Endocrinology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; ²Division of Dialysis and Apheresis, The University of Tokyo Hospital, Tokyo, Japan; and ³Department of Acute Medicine, The University of Tokyo Hospital, Tokyo, Japan

Introduction: For patients with end-stage renal disease (ESRD), due to the heterogeneity of the population, appropriate risk assessment approaches and strategies for further follow-up remain scarce. We aimed to conduct a pilot study for better risk stratification, applying machine learning–based classification to patients with ESRD who newly started maintenance hemodialysis.

Methods: We prospectively studied 101 patients with ESRD, who were new to maintenance hemodialysis therapy, between August 2016 and March 2018. Baseline values of variables such as blood and urine tests were obtained before the initiation of hemodialysis. Agglomerative hierarchical clustering was conducted with the collected continuous data. The resulting clusters were followed up for the primary outcome of 1-year mortality, as analyzed by the Kaplan-Meier survival curve with log-rank test and the Cox proportional hazard model.

Results: The participants were divided into 3 clusters (cluster 1, $n = 62$; cluster 2, $n = 15$; cluster 3, $n = 24$) by hierarchical clustering, using 46 clinical variables. Patients in cluster 3 showed lower systolic blood pressures, and lower serum creatinine and urinary liver-type fatty acid-binding protein levels, before the initiation of hemodialysis. Consequently, cluster 3 was associated with the highest 1-year mortality in the study cohort ($P < 0.001$), and the difference was significant after adjustment for age and sex (hazard ratio: 10.2; 95% confidence interval: 2.94–46.8, cluster 1 as reference).

Conclusion: In this proof-of-concept study, hierarchical clustering discovered a subgroup with a higher 1-year mortality at the initiation of hemodialysis. Applying machine learning–derived classification to patients with ESRD may contribute to better risk stratification.

Kidney Int Rep (2020) 5, 1188–1195; <https://doi.org/10.1016/j.ekir.2020.05.007>

KEYWORDS: end-stage renal disease; hemodialysis; hierarchical clustering; machine learning; risk prediction; renal replacement therapy

© 2020 International Society of Nephrology. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

A considerable number of patients with chronic kidney disease worldwide require initiation of maintenance dialysis therapy, despite every effort to prevent chronic kidney disease progression¹; the number of patients newly registered as having ESRD has continued increasing.^{2,3} The patients starting dialysis form a heterogeneous population; their incidences of underlying renal diseases, speeds of declining residual renal function, responsiveness to

treatment, and comorbidities vary greatly. It has become increasingly important to address this heterogeneity with precise and personalized care, to ameliorate the health and economic burdens of ESRD.⁴

Given the markedly higher mortality rate of patients with ESRD with dialysis when compared with that of the general population,³ risk evaluation, for effective intervention strategy, is warranted; this is especially true at the initiation of dialysis. One study reported the most vulnerable period, with the highest mortality, to be the first several months of dialysis therapy.⁵ Several previous studies have suggested some clinical risk factors to help inform predictions of maintenance dialysis patients' prognoses; lower systolic blood pressure (<140 mm Hg)⁶ or higher pH (≥ 7.40)⁷ was associated with higher mortality. However, there is a

Correspondence: Kent Doi, Department of Acute Medicine, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo, Tokyo, 1138655, Japan. E-mail: kdoi-tky@umin.ac.jp

⁴YK and TY contributed equally to this work.

Received 10 January 2020; revised 5 May 2020; accepted 11 May 2020; published online 23 May 2020

paucity of evidence supporting the introduction of multidimensional data analysis for risk stratification in patients with ESRD.

Recently, some researchers have applied machine learning techniques for risk stratification in clinical medicine⁸ and have succeeded in refining optimal treatment approaches.⁹ Machine learning contains an advantage in discovering similarities within multidimensional datasets, as given by particular protocols. With unsupervised learning, a subcategory of machine learning, past studies have succeeded in identifying intrinsic subgroups within heterogeneous populations, such as patients with heart failure⁸ and those with obstructive pulmonary diseases.¹⁰ In contrast, the impact of applying machine learning to the ESRD population remains to be understood.

We hypothesized that a machine learning–derived technique would be useful for the risk stratification of patients on maintenance hemodialysis. This proof-of-concept study aimed to conduct a clustering analysis of patients with ESRD at the initiation of their hemodialysis, following them up to ascertain any associations between the resulting clusters and their clinical outcomes.

METHODS

Study Design and Population

We performed a prospective observational cohort study at the University of Tokyo Hospital, a tertiary general medical center. We recruited patients with ESRD who had started maintenance hemodialysis therapy between August 1, 2016, and March 31, 2018. All patients were hospitalized to start their hemodialysis during the observational period; the timing of dialysis initiation for each patient, as well as its modality, was decided by at least 2 nephrologists, independently of the present study. We excluded those who were younger than 20 years, those who withdrew from dialysis therapy within 3 months of initiation, those who were without laboratory data, either of blood or urine, at the dialysis initiation, or those who were without informed consent. Body weight, height, and blood pressure were all measured from all subjects on admission, and blood and urine samples were collected just before their first dialysis sessions. The list of collected data is provided in [Supplementary Table S1](#).

We conducted the study in accordance with the tenets of the Declaration of Helsinki; written informed consent was obtained from all participants or their next of kin. The study protocol was approved by the Clinical Research Review Board of The University of Tokyo (study ID: 11239).

Clinical Outcomes

The primary outcome was 1-year mortality after the initiation of hemodialysis therapy. The secondary outcome was the duration of hospital stay for those who were discharged from the hospital alive. We performed the follow-ups on an outpatient clinic basis, using telephone calls for patients who did not visit after their discharge.

Statistical Analysis

First, the baseline characteristics of the continuous variables were summarized as medians with interquartile ranges, and the categorical variables were summarized as counts and percentages. We then applied agglomerative hierarchical clustering^{11,12} to the study cohort. Continuous variables with less than 10% of their data missing showed candidates for the clustering analysis; all the variables were standardized into normal distributions, and skewed variables were transformed onto logarithmic scales. The remaining missing data were imputed according to multivariate normal distribution model.¹³ To reduce redundant variables for clustering, we calculated Pearson's correlation coefficients among the candidate variables; either of 2 variables with correlation coefficients more than 0.6 was deleted from the clustering model, according to a previous article.⁸

Agglomerative hierarchical clustering is a technique that sequentially joins the 2 nodes of data with the shortest distance until all nodes are connected as one cluster. We adopted Ward's method for this analysis; the definition of the distance between 2 nodes was a weighted Euclidean distance, and the sum of within-cluster variance was controlled to be the minimum at each step.¹⁴ More detailed method for the clustering process is available in the [Supplementary Methods](#). The same clustering was also applied for intervariable analysis afterward, generating a visual heat map. We had to decide the optimal number of clusters, because agglomerative hierarchical clustering does not indicate the best trimming height for the resulting dendrogram. For this purpose, the cubic clustering criterion,¹⁵ C-index,¹⁶ and SD-index¹⁶ were validated for potential clusters numbers of 2 to 8. We also conducted K-means clustering analysis,¹⁷ another commonly used clustering method, in the same dataset for comparison. Detailed method for K-means clustering is available in the [Supplementary Methods](#).

The baseline characteristics, laboratory data, and clinical outcomes were compared among the divided clusters using Mann-Whitney *U* test for continuous variables and by Pearson's χ^2 test for categorical variables. Comparisons between more than 2 clusters were made through the Kruskal-Wallis test followed by the

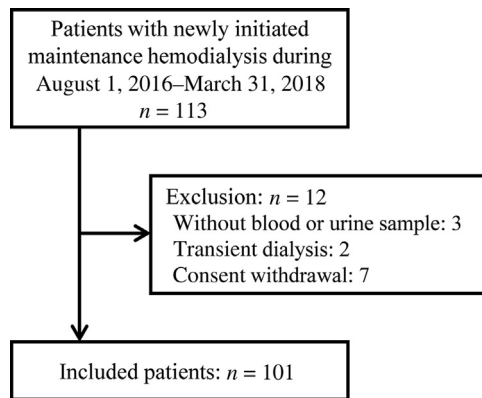


Figure 1. Patient flowchart.

Steel-Dwass test, when appropriate. The patients were subsequently followed up after the initiation of hemodialysis therapy, and mortality within clusters was compared using the Kaplan-Meier curve with log-rank test. The Cox proportional hazards model was used to assess the effect of clustering on mortality. Hospital stays following the first dialysis dates also were investigated; those who died without discharge and those who transferred to another hospital were excluded from the length of hospital stay analysis. As sensitivity analysis, 2 additional analyses were conducted. First, we performed agglomerative hierarchical clustering only with variables without missing value and no imputation. Second, in the survival analysis, systolic blood pressure, serum creatinine, serum potassium and B-type natriuretic peptide (BNP) were added as explanatory variables to the base model of the Cox proportional hazard model. We chose these variables based on previous literature⁶ and main reasons for introducing hemodialysis in patients with ESRD. To validate potential multi-collinearity in the models, variance inflation factor¹⁸ was assessed.

We used JMP Pro software (version 14.0.0; SAS Institute, Cary, NC) and “Nbclust” package from R software (version 3.5.0, R Foundation, Vienna, Austria)¹⁶ for validating the optimal number of clusters.

RESULTS

Study Population and Variable Selection

During the observational period, 113 patients with ESRD started maintenance hemodialysis therapy. With 12 patients meeting the exclusion criteria, we finally came up with 101 patients in the present study (Figure 1). Baseline demographic, physical, and laboratory data were collected, and coefficients of correlation for all possible pairs of the clinical continuous variables were obtained. We removed potentially correlated variables with coefficients of correlation

Table 1. Variables included in the hierarchical analysis

Domain	Variable (percentage of missing value, if not zero)
Demographic and physical characteristics	Age, BMI (2.0%), systolic blood pressure before hemodialysis
Blood test	Complete blood count (white blood cell, neutrophil percentage, monocyte percentage, eosinophil percentage, basophil percentage, hemoglobin, platelet, reticulocyte, [‡] MCV, MCHC), total protein, ALT, [‡] γ -GTP, [‡] uric acid, blood urea nitrogen, creatinine, sodium, potassium, calcium, phosphate, magnesium, glucose, HbA1c (1.0%), total cholesterol (1.0%), triglyceride (1.0%), [‡] high-density lipoprotein (2.0%), iron, [‡] ferritin, [‡] UIBC, CRP, [‡] BNP, [‡] iPTH, [‡] β 2-microglobulin (1.0%)
Urine test	pH, sodium, calcium (4.0%), urea nitrogen, creatinine, uric acid (1.0%), protein, [‡] NAG, [‡] α 1-microglobulin, L-FABP [‡] (2.0%)

ALT, alanine transaminase; BMI, body mass index; BNP, B-type natriuretic peptide; CRP, C-reactive protein; γ -GTP, γ -glutamyl transpeptidase; HbA1c, hemoglobin A1c; iPTH, intact parathyroid hormone; L-FABP, liver-type fatty acid-binding protein; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; NAG, N-acetyl- β -D-glucosaminidase; UIBC, unsaturated iron binding capacity.

[‡]Logarithmic transformation was conducted before clustering because of skewed distribution.

larger than 0.6 from further analysis. Urine volume was not adopted for clustering because of a relatively high rate of missing values (12.9%). The 46 variables finally adopted by the clustering analysis and percentages of missing values are shown in Table 1.

Clustering and the Comparison of Patients in 3 Clusters

Agglomerative hierarchical clustering in the 101 included patients was performed using these 46 variables. Clustering in variables was also conducted, and both sets of results are presented as a mapping graphic in Figure 2; each row represents 1 patient, and each column represents 1 clinical variable. The heat map shows several red zones, indicating relatively high values, especially in upper left and lower right area. The upper left area contains high values clinically associated with diabetes (blood glucose, hemoglobin A1c), hypertension (systolic blood pressure), dyslipidemia (cholesterol, triglycerides), and chronic kidney disease (urinary protein, urinary pH, urinary liver-type fatty acid-binding protein). In contrast, the lower right area contains some markers associated with inflammation (white blood cell, neutrophil, C-reactive protein) and fluid overload (BNP).

An optimal number of clusters were validated by 3 different indices. Using the cubic clustering criterion produced a negative value and showed a monotonic decreasing trend, suggesting that the cohort was of a unimodal nature (Supplementary Figure S1).¹⁵ We subsequently calculated the C- and SD-indices, for which the minimum values generally indicate the optimal number of clusters. As shown in Supplementary Figure S1, local dips were observed with a cluster number of 3.

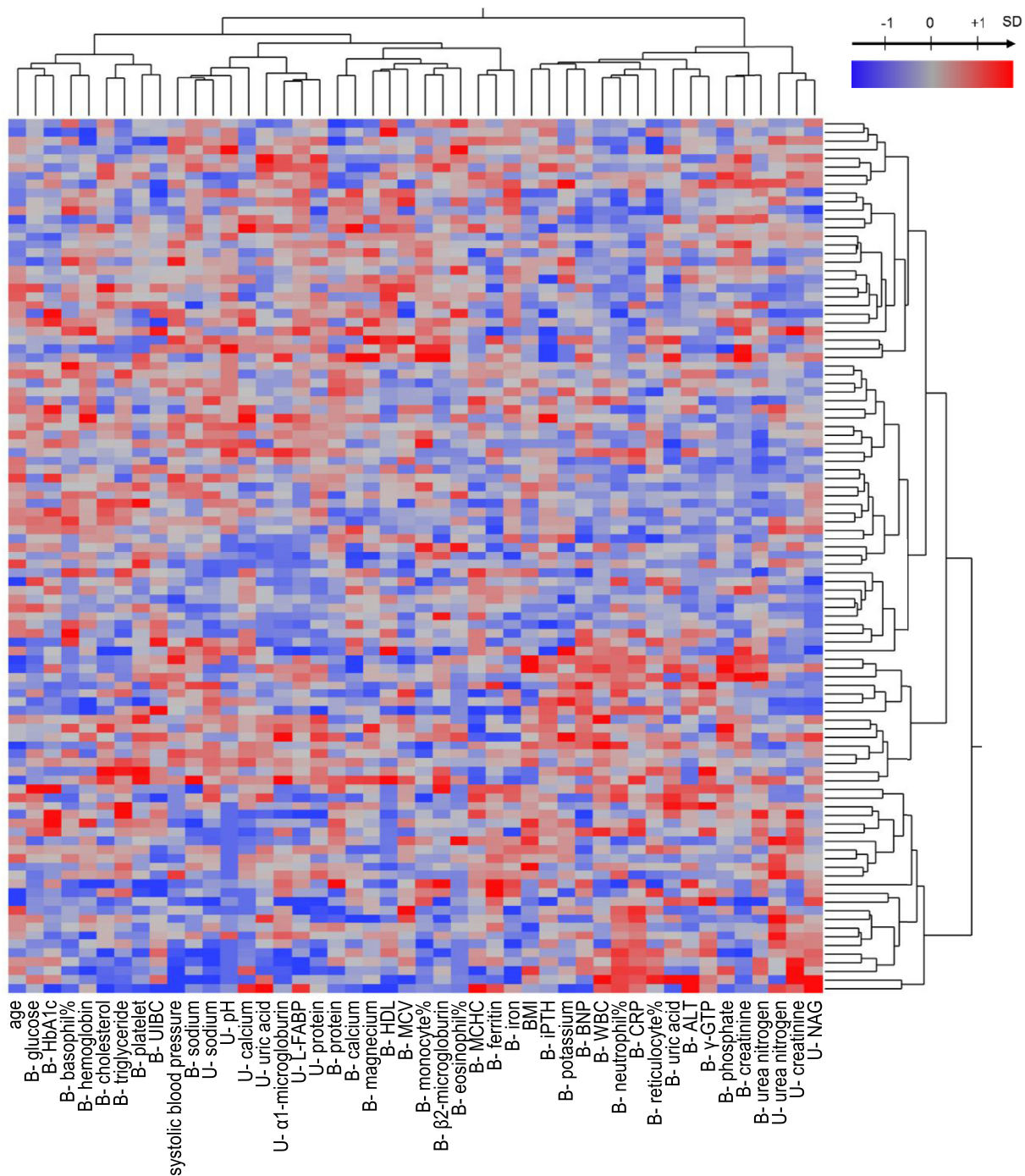


Figure 2. Hierarchical clustering. Based on 46 standardized clinical variables, agglomerative hierarchical clustering was performed in 101 patients who had recently started hemodialysis. Each row represents 1 patient, and each column represents 1 clinical variable, the name of which is shown in the bottom; prefix “B-” means data obtained from blood sample, and “U-” means data from urine sample. Hierarchical clustering of both rows and columns yielded a heat map, in which colors red and blue reflect comparatively high and low value scaled by SD, respectively. ALT, alanine transaminase; BMI, body mass index; BNP, B-type natriuretic peptide; CRP, C-reactive protein; γ -GTP, γ -glutamyl transpeptidase; HbA1c, hemoglobin A1c; iPTH, intact parathyroid hormone; L-FABP, liver-type fatty acid-binding protein; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; NAG, N-acetyl- β -D-glucosaminidase; SD, standard deviation; UIBC, unsaturated iron binding capacity

According to the aforementioned results of C- and SD-indexing, we cut the dendrogram of patients at the height that generated 3 clusters: 62 patients in cluster 1, 15 patients in cluster 2, and 24 patients in cluster 3. The baseline characteristics, underlying renal disease

incidence, and laboratory data for each cluster are shown in [Table 2](#). Predialysis systolic blood pressure, serum creatinine, and urinary liver-type fatty acid-binding protein were significantly higher in clusters 1 and 2, when compared with those of cluster 3. Higher

Table 2. Comparison of the 3 clusters

Variables	Total cohort (n = 101)	Cluster 1 (n = 62)	Cluster 2 (n = 15)	Cluster 3 (n = 24)	P value
Age, yr	67.0 [52.5–76.0]	70.5 [54.0–77.3] ^a	54.0 [43.0–64.0] ^a	67.5 [52.8–75.0]	0.006 ^b
Female, n (%)	34 (33.7)	20 (32.3)	10 (66.7)	4 (16.7)	0.005 ^b
Height, cm	163 [154–169]	162 [154–169]	154 [151–168]	166 [160–170]	0.093
Body weight, kg	62.6 [52.1–74.6]	60.9 [54.1–71.8]	66.5 [42.3–78.0]	69.6 [52.4–87.2]	0.28
Systolic blood pressure, mm Hg	150 [130–163]	150 [140–170] ^a	160 [140–175] ^c	130 [120–140] ^{a,c}	< 0.001 ^b
Underlying renal disease, n (%)					0.21
Diabetic kidney disease	36 (35.6)	22 (35.5)	6 (40.0)	8 (33.3)	
Nephrosclerosis	14 (13.9)	9 (14.5)	3 (20.0)	2 (8.3)	
Chronic glomerulonephritis	11 (10.9)	7 (11.3)	2 (13.3)	2 (8.3)	
ADPKD	7 (6.9)	7 (11.3)	0 (0)	0 (0)	
IgA nephropathy	4 (4.0)	4 (6.5)	0 (0)	0 (0)	
Others	29 (28.7)	13 (21.0)	4 (26.7)	12 (50.0)	
Blood test					
Total protein, g/dl	6.0 [5.5–6.5]	6.2 [5.5–6.5]	5.8 [5.4–6.2]	6.2 [5.5–6.7]	0.53
Blood urea nitrogen, mg/dl	92.7 [77.0–108.4]	91.6 [75.6–102.1]	90.5 [78.7–119.4]	102.0 [86.0–118.5]	0.077
Creatinine, mg/dl	8.49 [6.98–10.97]	8.89 [7.28–11.70] ^a	10.81 [7.84–12.66] ^c	6.74 [5.92–8.65] ^c	< 0.00 ^b
Potassium, mmol/l	4.2 [3.8–4.8]	4.1 [3.8–4.5] ^a	4.9 [4.4–5.3] ^{a,c}	4.3 [3.8–4.7] ^c	0.001 ^b
White blood cell, × 10 ³ /mm ³	5.9 [4.9–8.4]	5.5 [4.6–6.1] ^{a,c}	9.3 [7.9–11.5] ^a	7.2 [5.4–9.4] ^c	< 0.00 ^b
Hemoglobin, g/dl	8.9 [7.9–9.6]	9.2 [8.4–9.9]	7.9 [7.0–9.3]	8.4 [7.5–9.3]	0.020 ^b
CRP, mg/dL	0.24 [0.09–1.85]	0.14 [0.05–0.34] ^{a,c}	0.75 [0.24–3.89] ^a	2.20 [0.41–5.64] ^c	< 0.001 ^b
Hemoglobin A1c, %	5.7 [5.3–6.2]	5.7 [5.3–6.3]	5.6 [5.1–5.9]	5.6 [5.3–6.4]	0.56
BNP, pg/ml	221 [76–954]	206 [80–480] ^a	1686 [613–2063] ^{a,c}	148 [43–1175] ^c	< 0.001 ^b
β2-microglobulin, mg/l	17.2 [14.5–20.5]	17.5 [14.9–20.1]	20.6 [17.6–23.6] ^a	14.9 [13.4–18.3] ^b	0.014 ^b
Urine test					
pH	6.0 [5.0–7.0]	6.5 [6.0–7.0] ^a	6.5 [5.5–7.0] ^c	5.0 [5.0–5.5] ^{a,c}	< 0.001 ^b
Creatinine, mg/dl	59.2 [44.8–78.4]	53.1 [45.7–74.2] ^a	50.8 [35.4–65.1] ^c	92.1 [56.5–112.8] ^{a,c}	0.002 ^b
Protein, mg/dl	167 [66.6–329]	212 [81.8–325] ^a	345 [139–510] ^c	75.0 [36.3–186] ^{a,c}	0.001 ^b
NAG, IU/L	7.3 [4.8–11.9]	6.8 [4.6–10.4] ^a	7.8 [4.6–15.5]	9.7 [7.1–13.6] ^a	0.005 ^b
α1-microglobulin, mg/l	52.2 [34.4–73.3]	51.3 [36.6–68.3]	69.1 [42.8–80.2]	48.3 [15.6–74.6]	0.23
L-FABP, ng/ml	77.6 [39.2–110.8]	89.0 [42.1–114.6] ^a	77.8 [64.1–119.7] ^c	46.7 [18.1–91.7] ^{a,c}	0.008 ^b
Urine volume, ml/d	1120 [743–1510]	1100 [740–1530]	1200 [630–2000]	1030 [748–1313]	0.77
Hospital stay, d	8 [5–20]	6 [5–11] ^{a,c}	20 [12–42] ^a	21 [7–48] ^c	< 0.001 ^b
90-d mortality, n (%)	6 (5.9)	1 (1.6)	0 (0)	5 (20.8)	0.002 ^b
1-yr mortality, n (%)	14 (13.9)	3 (4.8)	2 (13.3)	9 (37.5)	< 0.001 ^b

ADPKD, autosomal dominant polycystic kidney disease; BNP, B-type natriuretic peptide; CRP, C-reactive protein; L-FABP, liver-type fatty acid-binding protein; N-acetyl-β-D-glucosaminidase.

^{a,c}Statistically significant difference ($P < 0.05$) in comparing 2 groups using the Steel-Dwass test for multiple comparison.

^b $P < 0.05$ by the Kruskal-Wallis test or Pearson's χ^2 test.

values of serum potassium and BNP were observed in cluster 2. Cluster 3 showed significantly higher serum C-reactive protein than cluster 1, whereas urinary pH and urinary protein excretion in cluster 3 remained the lowest of all 3 clusters. In terms of underlying renal disease, there was not a significant difference in the cluster percentages of diabetic kidney disease and nephrosclerosis; however, cluster 1 included all the patients with autosomal dominant polycystic kidney disease and IgA nephropathy. In K-means clustering, most patients in original cluster 1 and 3 remained, but patients in cluster 2 were divided into new cluster 1 and 3 (Supplementary Figure S2).

Prognosis of the Patients in Each Cluster

Follow-up was carried out for all participants for at least 1 year. There was a significant difference in overall survival among the 3 clusters ($P < 0.001$, Figure 3), with the difference between cluster 3 and

cluster 1 still significant after Bonferroni's correction for multiple comparisons ($P < 0.001$). The age- and sex-adjusted hazard ratio of mortality, for cluster 3 over cluster 1, was 10.2 (95% confidence interval: 2.94–46.8; $P < 0.001$). Hospital stay duration, since initiation of dialysis, was significantly longer in cluster 2 (20 [12–42] days) and cluster 3 (21 [7–48] days), when compared with cluster 1 (6 [5–11] days; $P < 0.001$ and $P = 0.014$ for cluster 2 vs. 1 and cluster 3 vs. 1, respectively; Supplementary Figure S3). Although some patients were reclassified into different clusters in K-means clustering, the survival difference in the resulting clusters by K-means was still significant ($P = 0.007$, Supplementary Figure S2).

Sensitivity Analysis

There were 15 missing values of 4646 observations (Table 1). With the remaining 37 variables without imputation, we again performed agglomerative

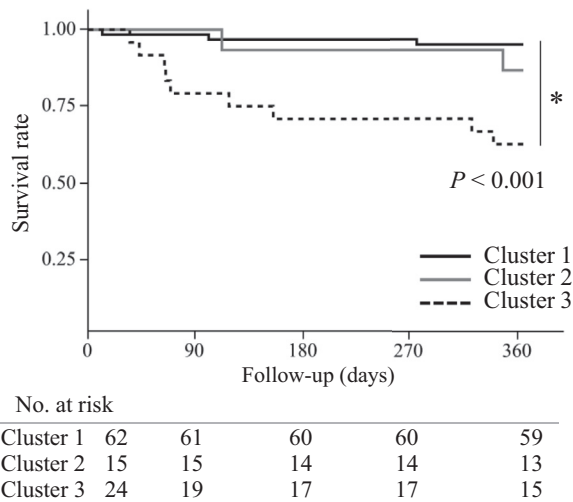


Figure 3. Survival analysis of the 3 clusters suggested by the hierarchical clustering. Patients in cluster 3 showed a significantly worse survival rate compared with cluster 1, in the Kaplan-Meier analysis of the follow-ups 1 year after hemodialysis therapy was initiated. The difference between cluster 3 and cluster 1 was still significant after Bonferroni's correction for multiple comparisons ($P < 0.001$). * $P < 0.05$.

hierarchical clustering. All patients in cluster 1 remained in their original cluster, although 8 and 12 patients in cluster 2 and 3 moved to the new cluster 1, respectively. Kaplan-Meier survival analysis based on the new classification still showed significant difference in the 3 clusters ($P < 0.001$, [Supplementary Figure S4](#)). Finally, we added potential covariates (systolic blood pressure, serum creatinine, serum potassium, and BNP) to the base model of the Cox proportional hazards model. Even in these additional models, the effect of clustering on 1-year mortality remained significant for the survival outcome ([Table 3](#)). Variance inflation factors for the resulting cluster group in each model, model 1–5 in [Table 3](#), were 1.12, 1.39, 1.28, 1.25, and 1.24, respectively. These values were not indicative of serious collinearity.

DISCUSSION

In this pilot study, agglomerative hierarchical clustering was applied to 101 patients with ESRD newly starting maintenance hemodialysis, with 46 standardized continuous variables per person. Validating the optimal number of clusters by several existing indices, we finally identified 3 clusters (cluster 1, $n = 62$; cluster 2, $n = 15$; cluster 3, $n = 24$) within this heterogeneous population. The 3 clusters differed in their clinical data. Patient data for 1 year showed a significantly lower survival rate within cluster 3 compared with those in cluster 1. This is also supported by the Cox proportional hazards model adjustment for age and sex. Although hospital stays after initiation of dialysis therapy were significantly longer in cluster 2 than cluster 1, the subsequent survival rates of the 2 groups were similar.

Table 3. Cox proportional hazard model for 1-year mortality in patients initiating hemodialysis

Variable	Model 1 HR [95% CI]	Model 2	Model 3	Model 4	Model 5
Age	1.07 ^a [1.01–1.13]	1.06 ^a [1.01–1.13]	1.07 ^a [1.02–1.13]	1.07 ^a [1.01–1.13]	1.07 ^a [1.01–1.14]
Sex: women	0.31 [0.04–1.43]	0.22 [0.03–1.14]	0.26 [0.03–1.18]	0.25 [0.03–1.33]	0.38 [0.05–1.51]
Cluster 3 (cluster 1 as reference)	10.2 ^b [2.94–46.8]	6.44 ^a [1.55–34.4]	8.02 ^b [2.19–38.2]	11.8 ^b [3.33–55.3]	10.5 ^b [3.02–48.3]
Systolic blood pressure		0.98 [0.95–1.01]			
Serum creatinine			0.76 ^a [0.57–0.99]		
Serum potassium				0.49 [0.20–1.18]	
BNP					1.00 [0.99–1.00]

BNP, B-type natriuretic peptide; CI, confidence interval; HR, hazard ratio.

^a $P < 0.05$.

^b $P < 0.01$.

Model 1 (base model): age + sex + cluster; model 2: model 1 + systolic blood pressure; model 3: model 1 + serum creatinine; model 4: model 1 + serum potassium; model 5: model 1 + BNP.

Importantly, the resulting clusters presented valuable clinical implications; physicians should be aware of the remarkably high mortality of the patients in cluster 3 and should provide rapid interventions to mitigate potentially modifiable risk factors, when required for the high-risk population. In the heat map of [Figure 2](#), cluster 1 corresponded to patient group with the upper left red zone, and cluster 3 corresponded to those with the lower right red zone. Cluster 2 contained both features of clusters 1 and 3, but hot spot on high BNP and serum potassium level was remarkable. We interpreted the individual cluster results as follows. The largest group, cluster 1, might be composed of patients with gradual progressions of chronic kidney disease, leading to the initiation of hemodialysis with fewer complications, requiring the shortest hospital stay. Characteristics of cluster 2, which include longer hospital stays than cluster 1, suggested that the accumulation of fluid and small molecules such as potassium and creatinine might have frequently occurred. Despite the longer hospitalizations, it was notable that the 1-year survival rate of cluster 2 was comparable to that of cluster 1. Patients in cluster 3 had features distinctive from the other 2 clusters and showed the worst mortality in the study cohort. It might be possible that these patients had higher levels of inflammation and other organ complications. This speculation was supported by the higher serum C-reactive protein levels and lower systolic blood pressures observed in cluster 3, both of which were reportedly associated with poorer outcome in patients with ESRD.^{6,19} Moreover, urine acidification

ability and barriers for proteinuria were rather preserved in cluster 3. Our study also confirmed an earlier observation: the 1-year follow-up of the 3 clusters revealed that most mortal events occurred in the first half of the year (Figure 3). This corresponded to the previous report based on an international survey.⁵

Clustering analysis is generally heuristic, and different clustering methods often generate different results.¹⁷ In our dataset, the clustering result by agglomerative hierarchical method was not consistent with the result by K-means method. Both clustering methods contain their strengths over the other. Hierarchical clustering offers comprehensive process of classification and the result allows us to speculate unknown shared features within each cluster. On the other hand, K-means method can analyze large sample size and is less susceptible to outlier.²⁰ In the present study, we selected agglomerative hierarchical clustering due to the sample size and interpretability; the comprehensive process of hierarchical clustering led to our discussion on the resulting clusters with Figure 2, but dataset with more outliers or larger sample size may be appropriate to the K-means method. It is noteworthy that some other machine learning algorithms, including neural network, also require a much larger sample size, and understanding the reasons behind their prediction is usually difficult.

Several recent studies have applied unsupervised machine learning–derived classification technique to reestablish some established disease definitions, such as heart failure⁸ and obstructive pulmonary diseases.¹⁰ Still others have suggested a valuable use of machine learning to be for the optimization of care for septic patients.⁹ These approaches have also been applied in other fields. Nephrology, via computer-based pathological evaluation of kidney biopsy images,^{21,22} artificial intelligence–based anemia management programs,²³ and the predictions of acute kidney injuries by deep learning approaches²⁴ have formed part of this innovative movement.

Unsupervised machine learning enables us to discover intrinsic patterns in multidimensional data²⁵; our subjective impression, otherwise, is easily affected by limited numbers of remarkable data; it is almost impossible to evaluate several clinical parameters at once. With increasing numbers of multidimensional clinical data available, due to the recent advent of electronic medical records, hierarchical clustering analysis has great potential for developing personalized therapies and realizing better patient care. In conventional analysis, in contrast, we must select important variables from a large dataset based on previous publication, experience, and clinical rationale.

To the best of our knowledge, the present study is the first to report a clinical application of hierarchical clustering to patients with ESRD. In addition, we performed

intervariable clustering for patients. This dual-direction clustering yielded a color map of the results (Figure 2), which visually represents the characteristics of the participants; this technique was recently named as “phenomapping.”^{8,26} We applied this method to the data of the patient newly starting hemodialysis to assist our visual perception of individual patients. Moreover, the clustering may indicate similarities or redundancies among the clinical variables and provide an optimal method of variable reduction. Thus, with further studies, it would be possible to select a core clinical dataset, obtained at the initiation of hemodialysis, for risk stratification of patients with ESRD.

We acknowledge several limitations. First, the present study was conducted on a single-center basis with a modest sample size, which may limit the generalizability of the results. Second, prognoses of the patients over a longer period than observed were unknown, because we predefined the follow-up plan to include the time period of highest risk for patients starting dialysis, according to the previous report.⁵ Third, the clinical management, including dialysis modality and medication after discharge, depended on each hemodialysis clinic. Fourth, other machine learning techniques may be better for larger populations; some preceding application of hierarchical clustering implied that several dozen to several hundred patient data is appropriate to this clustering method,^{8,11} and it may not be applicable to analyze a much larger sample size. Further study is required to validate the results of our study in external dataset, and investigating the optimal clustering method for a larger dataset.

In conclusion, we have demonstrated a proof of concept that agglomerative hierarchical clustering, an unsupervised machine learning technique, can be applied to a population of patients newly starting maintenance hemodialysis. Using the intrinsic nature of the methodology, we found that the resulting classification was associated with 1-year mortality and length of hospital stay. With future investigations, clinical application of this technique may give a better understanding for risk stratification approaches to support patients with ESRD about to start hemodialysis.

DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

We express our appreciation to the clinical staff in the Division of Hemodialysis and Apheresis at the University of Tokyo Hospital for their dedicated patient care and sample collection. The authors thank Enago (www.enago.jp) for the English language review.

AUTHOR CONTRIBUTIONS

TY and YH conceived and designed the study. YK and TY recruited the participants, collected, analyzed, and interpreted the data. KD supervised the data analyses and interpretation of the results. YK drafted the manuscript, and MN and KD critically revised it. All authors reviewed and approved the final version of this manuscript.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Supplementary Methods. Detailed method of clustering analysis.

Table S1. Data list obtained from the participants.

Figure S1. Validation of the optimal number of clusters.

Figure S2. Clustering analysis by K-means method.

Figure S3. Length of hospital stay since dialysis therapy was initiated.

Figure S4. Agglomerative hierarchical clustering with 37 variables without missing data.

STROBE statement

REFERENCES

1. Thomas B, Wulf S, Bikbov B, et al. Maintenance dialysis throughout the world in years 1990 and 2010. *J Am Soc Nephrol.* 2015;26:2621–2633.
2. Nitta K, Masakane I, Hanafusa N, et al. 2017 Annual Dialysis Data Report, JSDT Renal Data Registry. *J Jpn Soc Dial Ther.* 2018;51:68.
3. Saran R, Robinson B, Abbott KC, et al. US Renal Data System 2018 Annual Data Report: epidemiology of kidney disease in the United States. *Am J Kidney Dis.* 2019;73:A7–A8.
4. Usvyat L, Dalrymple LS, Maddux FW. Using technology to inform and deliver precise personalized care to patients with end-stage kidney disease. *Semin Nephrol.* 2018;38:418–425.
5. Robinson BM, Zhang J, Morgenstern H, et al. Worldwide, mortality risk is high soon after initiation of hemodialysis. *Kidney Int.* 2014;85:158–165.
6. Sumida K, Molnar MZ, Potukuchi PK, et al. Blood pressure before initiation of maintenance dialysis and subsequent mortality. *Am J Kidney Dis.* 2017;70:207–217.
7. Yamamoto T, Shoji S, Yamakawa T, et al. Predialysis and postdialysis pH and bicarbonate and risk of all-cause and cardiovascular mortality in long-term hemodialysis patients. *Am J Kidney Dis.* 2015;66:469–478.
8. Shah SJ, Katz DH, Selvaraj S, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation.* 2015;131:269–279.
9. Komorowski M, Celi LA, Badawi O, et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med.* 2018;24:1716–1720.
10. Hirai K, Shirai T, Suzuki M, et al. A clustering approach to identify and characterize the asthma and chronic obstructive pulmonary disease overlap phenotype. *Clin Exp Allergy.* 2017;47:1374–1382.
11. Moore WC, Meyers DA, Wenzel SE, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med.* 2010;181:315–323.
12. Gagnon P, Casaburi R, Saey D, et al. Cluster analysis in patients with GOLD 1 chronic obstructive pulmonary disease. *PLoS One.* 2015;10:e0123626.
13. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol.* 2010;171:624–632.
14. Hirano S, Sun X, Tsumoto S. Comparison of clustering methods for clinical databases. *Information Sciences.* 2004;159:155–165.
15. D'Amours R. *SAS Technical Report A-108, Cubic Clustering Criterion.* Cary, NC: SAS Institute Inc; 1998.
16. Charrad M, Ghazzali N, Boiteux V, et al. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw.* 2014;61:1–36.
17. Steinbach M, Karypis G, Kumar V. A comparison of document clustering techniques. Paper presented at: TextMining Workshop at the Sixth International Conference on Knowledge Discovery and Data Mining. August 20–23, 2000; Boston, MA.
18. Zheng X, Zhang R, Liu X, et al. Association between cumulative exposure to ideal cardiovascular health and arterial stiffness. *Atherosclerosis.* 2017;260:56–62.
19. Iseki K, Tozawa M, Yoshi S, et al. Serum C-reactive protein (CRP) and risk of death in chronic dialysis patients. *Nephrol Dial Transplant.* 1999;14:1956–1960.
20. Liao M, Li Y, Kianifard F, et al. Cluster analysis and its application to healthcare claims data: a study of end-stage renal disease patients who initiated hemodialysis. *BMC Nephrol.* 2016;17:25.
21. Barisoni L, Nast CC, Jennette JC, et al. Digital pathology evaluation in the multicenter Nephrotic Syndrome Study Network (NEPTUNE). *Clin J Am Soc Nephrol.* 2013;8:1449–1459.
22. Barros GO, Navarro B, Duarte A, et al. PathoSpotter-K: a computational tool for the automatic identification of glomerular lesions in histological images of kidneys. *Sci Rep.* 2017;7:46769.
23. Barbieri C, Molina M, Ponce P, et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. *Kidney Int.* 2016;90:422–429.
24. Tomasev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature.* 2019;572:116–119.
25. Shah P, Kendall F, Khozin S, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *Digital Med.* 2019;2:69.
26. Bakir M, Jackson NJ, Han SX, et al. Clinical phenomapping and outcomes after heart transplantation. *J Heart Lung Transplant.* 2018;37:956–966.