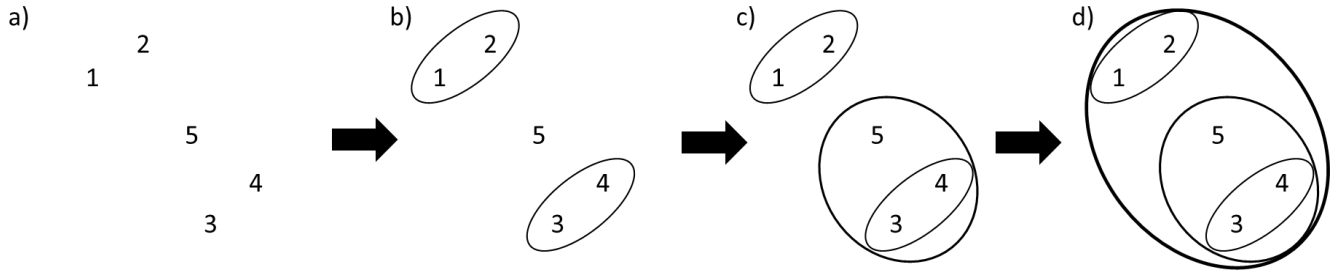
Supplementary Methods

Clustering analysis is a method to perform grouping of similar objects into several clusters. Ideal clustering yields a set of clusters in which objects from different clusters are distinct from each other, while objects in the same cluster are similar. Generally, clustering analysis is aimed to find hidden pattern(s) in dataset.

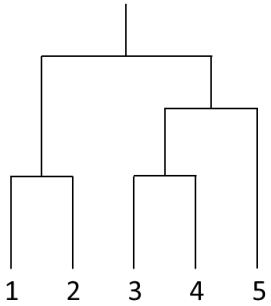
Agglomerative hierarchical clustering: At the beginning of agglomerative hierarchical clustering, each object exists as a single cluster (**Panel 1a**). Based on a definition of distance between objects, the nearest clusters are sequentially merged until generating one big cluster (**Panel 1b, c, d**). This approach is also expressed as dendrogram in **Panel 2**. In our study, each patient contained 46 clinical parameters, which corresponded to 46-dimentional object. In creating a new set of clusters, sum of squared distances within suggested clusters were calculated, and new set of clusters were determined to minimize the sum value. This method is known as Ward's method. The distance between two objects was defined as Euclidean distance as follows:

$$Euclidean\ distance\ between\ \boldsymbol{a}\ \text{and}\ \boldsymbol{b} = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2} = \sqrt{\sum_{k=1}^{n}(b_k - a_k)^2}$$

* $\boldsymbol{a}$ and $\boldsymbol{b}$ are n-dimentional objects; a_x, b_x: xth element of each object



Panel 1: process of agglomerative hierarchical clustering



Panel 2: dendrogram generated as a result of agglomerative hierarchical clustering shown in panel 1

K-means clustering: The K-means method is a non-hierarchical clustering algorithm. First, number of clusters ("k") should be determined. K points are then randomly selected as center of each cluster ("centroid"). Other observations are assigned to their nearest centroids to form k clusters. When all objects have been assigned, the positions of the k centroids are moved to mean points of each cluster. Calculation and reclassification are repeated until the centroids no longer move.

Reference:

1. 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.

2. Karypis, M. S. G., Kumar, V., & Steinbach, M. (2000, May). A comparison of document clustering techniques. In TextMining Workshop at KDD2000 (May 2000).

Table S1. Data obtained from the participants

Domain	Variable
Basic characteristics	Age, sex, underlying renal disease
Physical data	Body weight, height, BMI, systolic and diastolic blood pressure*, urine volume
Blood test	Complete blood count, total protein, albumin, AST, ALT, γ-GTP, ALP, uric acid, blood urea nitrogen, creatinine, sodium, potassium, chloride, calcium, phosphate, magnesium, CRP, total cholesterol, triglyceride, high-density lipoprotein, iron, ferritin, unsaturated iron binding capacity, BNP, iPTH, glucose, HbA1c, β2-microglobulin
Urinary test	Specific gravity, pH, sodium, potassium, chloride, calcium, phosphate, magnesium, urea nitrogen, creatinine, uric acid, protein, NAG, α1-microglbulin, L-FABP

*, Blood pressure was measured at the beginning of first dialysis therapy.

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; 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; NAG, N-acetyl-β-D-glucosaminidase

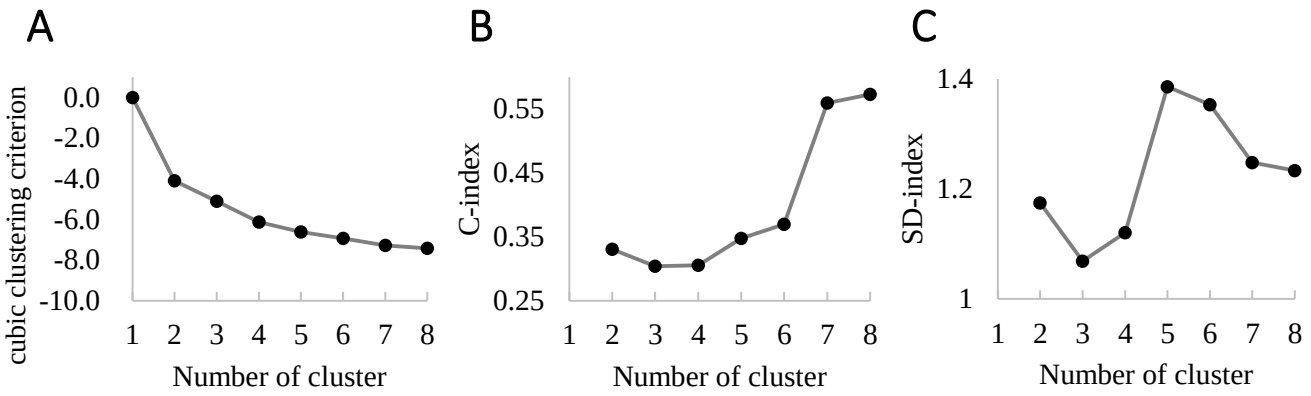


Figure S1: Validation of the optimal number of clusters.
The cubic clustering criterion, C-index, and SD-index were calculated in order to determine the best number of clusters for the 101 patients. The maximum value given by the cubic clustering criterion and the minimum values obtained using the C-index and SD-index are generally considered to give the optimal number of clusters. (A) The cubic clustering criterion shows a monotonic, decreasing trend, suggesting a unimodal nature of the cohort. (B, C) Both C-index and SD-index reach their minima with cluster three. This value has been adopted for subsequent analyses.

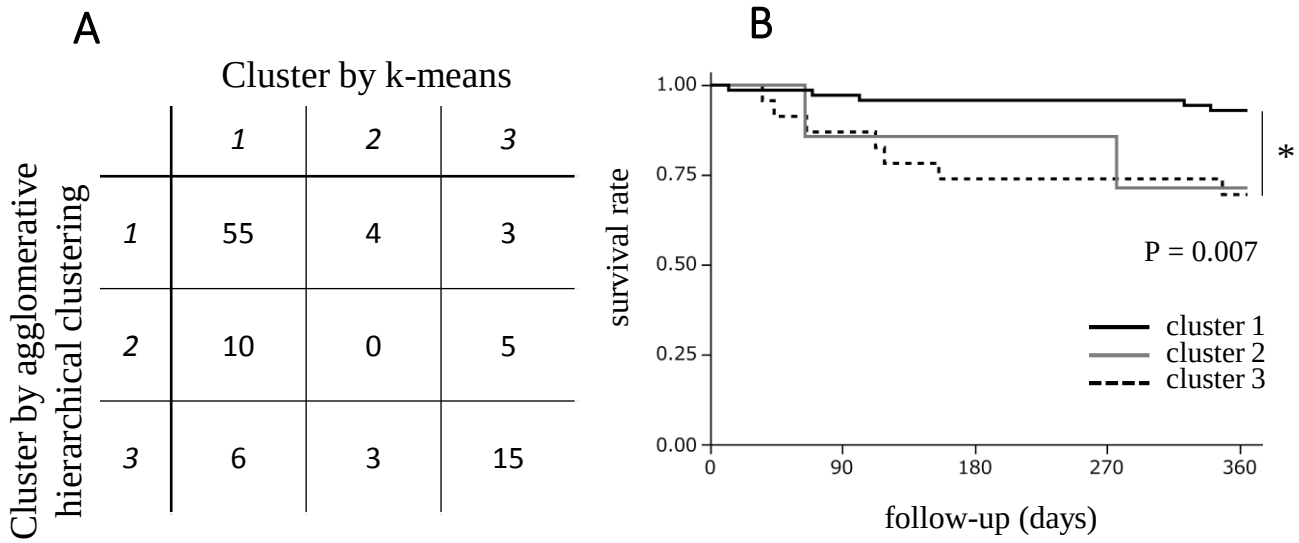


Figure S2: Clustering analysis by k-means method.
(A) Reclassification by k-means method generated different clustering result, compared with the original agglomerative hierarchical clustering approach. Most patients in the original cluster 1 and 3 remained, but patients in the cluster 2 were divided into new cluster 1 and 3.
(B) Kaplan-Meier survival analysis still showed significant difference in the new clusters (P=0.007). *, P < 0.05

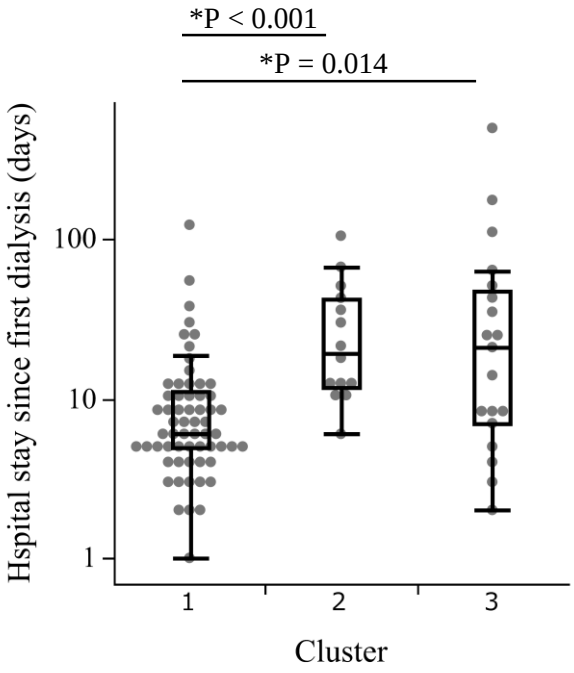


Figure S3: Length of hospital stay since dialysis therapy was initiated. Length of hospital stay since the first hemodialysis therapy was compared for the three clusters. Cluster 1 (6 [5, 11] days) showed significantly shorter hospital stay compared with cluster 2 (20 [12, 42] days) and cluster 3 (21 [7, 48] days). *, P < 0.05

A

Clustering with 37 variables without imputation

Clustering with 46 variables (original cluster)	1	2	3
1	62	0	0
2	8	7	0
3	12	0	12

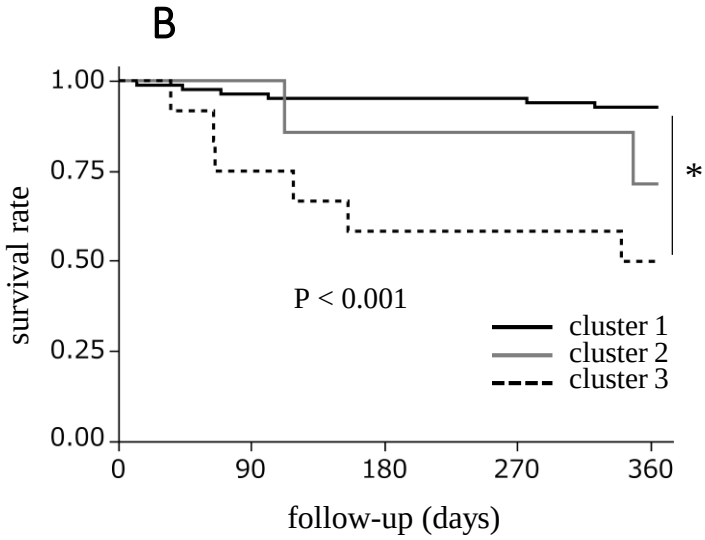


Figure S4: Agglomerative hierarchical clustering with 37 variables without missing data. As sensitivity analysis, agglomerative hierarchical clustering was performed in 101 patients only with 37 variables without missing value. (A) All patients in cluster 1 remained, but 8 and 12 patients in original cluster 2 and 3 moved to new cluster 1, respectively. (B) Kaplan Meier survival analysis still showed significant difference in the new clusters ($P < 0.001$). *, $P < 0.05$

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	6-7
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-8
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	7-8
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, explain how loss to follow-up was addressed	7
		(*e*) Describe any sensitivity analyses	9
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	9
		(c) Consider use of a flow diagram	Fig1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9, Table 2
		(b) Indicate number of participants with missing data for each variable of interest	9, Table 1, Table S1
		(c) Summarise follow-up time (eg, average and total amount)	11
Outcome data	15*	Report numbers of outcome events or summary measures over time	11, Figure 3

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (*b*) Report category boundaries when continuous variables were categorized (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	12-13 Table 3 N/A N/A
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	14
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	17, 18
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15, 16
Generalisability	21	Discuss the generalisability (external validity) of the study results	17, 18
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	1 (None)

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.