SUPPLEMENTAL MATERIAL

Table S1. Baseline characteristics of the included and excluded patients

Variable	Included n=1996	Excluded n=612	P
Age, y	62.7±10.9	62.0±11.8	0.290
Female, n (%)	656 (32.9)	189 (30.9)	0.359
Smoker, n (%)	860 (43.1)	264 (43.3)	0.948
Drinker, n (%)	745 (37.4)	230 (37.7)	0.125
NIHSS on admission, median (IQR)	2 (1, 4)	2 (1, 4)	0.843
Medical history, n (%)			
Stroke	469 (23.5)	149 (24.4)	0.666
Hypertension	1403 (70.3)	434 (70.9)	0.767
Diabetes mellitus	433 (21.7)	135 (22.1)	0.848
Dyslipidemia	209 (10.5)	56 (9.2)	0.344
Coronary heart disease	214 (10.7)	77 (12.6)	0.201
Ischemic stroke classification, n (%)			<0.001
Large-artery atherosclerosis	1008 (57.8)	350 (62.9)	
Cardioembolism	55 (3.2)	34 (5.6)	
Small-artery occlusion	628 (36.0)	148 (26.6)	
Other	52 (3.0)	24 (4.3)	
History of Medication, n (%)			
Antiplatelet therapy	409 (20.5)	135 (22.1)	0.686
Statin therapy	201 (10.1)	75 (12.3)	0.299
Antihypertension therapy	1087 (54.5)	329 (53.8)	0.882
Medication at discharge, n (%)			
Antiplatelet therapy	1868 (93.6)	566 (92.5)	0.339
Statin therapy	1670 (83.7)	526 (85.9)	0.176
Antihypertension therapy	1334 (66.8)	397 (64.9)	0.368

IS: ischemic stroke; NIHSS: National Institute of Health stroke scale; IQR: interquartile range

S.2 Profiles Grouping Methods

S.2.1 Preprocessing of ABP data

We focused on one 24-hour cycle SBP measurement, which starts from 10:00 am to 10:00 am the next day. Furthermore, to align the SBP measurements across the observations, we construct a 15 minutes spaced grid $t_1 \cdots t_T$ on the time axis, where the 15 minutes gap was selected because the SBP measurements were taken roughly every 15 minutes during the day. We took the average of the measurements within the jth grid cell to represent the SBP level at time t_j.

Let $X_i(t_j)$ be the independent identically distributed SBP measurement at time t_j, $i = 1, \cdots, n$. Because of the irregular measurement times, patients may have missing measurements in some intervals. To indicate the missing measurement, we define $R_{ij} = 1$ if the patient i measures his/her blood pressure at t_j, $R_{ij} = 0$ otherwise. This missing mechanism is not disease relevant so that the missing at random assumption is appropriate in our study.

S.2.2 Functional clustering algorithms

Let C_i denote the classification of individual i and $C = (C_1, \cdots, C_n)'$, we model $X_i(t) = \sum_{c=1}^{M} I(C_i = c) \ _c(t) + f_i(t) + \ _i$, where $\ _c(t) = E\{X_i(t) \mid C_i = c\}$ is the mean function for the class c, $c = 1, \cdots, M$, and $\ _i$ is the mean 0 random error, f_i is an unknown function which describes the within class fluctuation for the patient i. We approximate $f_i(t)$ in the form of $P(t)' \ _i$, where P represents the d dimensional Fourier basis. We then estimate $\ _i$, $\ _c(t)$ by minimizing the following sum of squared error (SSE), $\sum_{i=1}^{n} \sum_{j=1}^{T} R_{ij} I(C_i = c) \{X_i(t_j) \ \ _c(t_j) \ \ P^T(t_j) \ _i\}^2$, for a given cluster

membership C, where $\gamma = (\gamma_1', \cdots, \gamma_n')'$.

When the cluster membership C is unknown, we implement the iterative clustering method of Chiou and Li[1]. Essentially, we iteratively minimize the SSE, alternating with respect to the cluster membership C, and the model parameters $_c$and . It is worth mentioning that this clustering method does not require regularly distributed measurement times, because it automatically handles cases with missing measurements. An R implementation of the clustering algorithm is available as the *funcy* package on CRAN.

REFERENCES

1. Chiou JM LP. Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society*. 2007;69:679-699