- 1 Outcome and life expectancy associated with novel subgroups of atrial fibrillation: a data-
- 2 driven cluster analysis
- 3
- 4 Yuefeng Yu¹, Jiang Li¹, Ying Sun¹, Bowei Yu¹, Xiao Tan^{2,3}, Bin Wang¹, Yingli Lu¹, Ningjian
- 5 $Wang^1$
- 6
- 7 Yuefeng Yu and Jiang Li contributed equally to this manuscript.
- 8
- 9 1 Institute and Department of Endocrinology and Metabolism, Shanghai Ninth People's
- 10 Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, China
- 11 2 School of Public Health, Zhejiang University, Hangzhou, China
- 12 3 Department of Medical Sciences, Uppsala University, Uppsala, Sweden
- 13
- 14 Short running title: reclassification of AF
- 15 Key words: atrial fibrillation; cluster analysis; data driven; risk profile; UK Biobank
- 16 Number of figures and tables: 5
- 17 Manuscript word count: 3062
- 18

19 Corresponding Author: Ningjian Wang, MD, PhD (he should receive the proofs and will

- 20 handle correspondence at all stages of refereeing and publication, also post-publication.)
- 21 Address: Institute and Department of Endocrinology and Metabolism, Shanghai Ninth
- 22 People's Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, 200011

- 23 China.
- 24 Telephone Number: 86-21-23271699
- 25 E-mail: wnj486@126.com
- 26
- 27 Co-corresponding Author: Yingli Lu, MD, PhD
- 28 Address: Institute and Department of Endocrinology and Metabolism, Shanghai Ninth
- 29 People's Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, 200011
- 30 China.
- 31 E-mail: luyingli2008@126.com
- 32

33 Abstract

34 Aim

35 This study aimed to develop a refined classification of atrial fibrillation (AF) according to36 comprehensive risk factors to identify patients at high-risk for poor prognosis and their life

- 37 expectancy.
- 38 Method

A total of 7,391 participants aged 40-69 with AF at baseline, from UK Biobank, were classified into five clusters, based on seven clustering variables including age and six risk factor categories (metabolic disease, respiratory disease, cardiovascular disease, renal/immune-mediated disease, mental health disease, acute illness). Difference in the risk of death and major complications, as well as reductions in life expectancy, among clusters were estimated. Replication was done in 2,399 participants with newly diagnosed AF within two years after baseline.

46 Results

47 Five distinct AF clusters were identified: acute illness-related, mental health-related, 48 cardiovascular disease-related, immune-and-renal disease-related, and respiratory-and-49 metabolic disease-related AF. Patients with respiratory-and-metabolic disease-related AF had 50 the highest risk of death, acute myocardial infarction, heart failure, and cerebral ischemic 51 stroke, while those with acute illness-related AF had the lowest corresponding risk. In 52 addition, compared with individuals with acute illness-related AF, those with respiratory-and-53 metabolic disease-related and mental health-related AF had the top 2 greatest loss of life 54 expectancy. Furthermore, genetic variants for AF had different effect among the five clusters.

55 Replication analysis confirmed the result stability.

56 Conclusion

- 57 A novel AF classification was developed, which provided insights into varying life
- 58 expectancy and risks of death and complications among AF subgroups with distinct
- 59 characteristics. It offers a practical approach for identifying high-risk patients, which might
- 60 help to tailor precise interventions for AF management.

62 Introduction

63	Atrial fibrillation (AF), as the most common sustained arrhythmia, imposes a huge burden on
64	individuals and public health worldwide ¹ . In the 2019 global statistics, there were
65	approximately 59.7 million people with AF and atrial flutter and the number is increasing 2 .
66	Even under existing treatment strategies, patients with AF still had higher risk of mortality,
67	heart failure, and stroke ³ . One possible explanation for these shortcomings is that the current
68	diagnosis of AF is mainly based on measurement of characteristics of AF episodes including
69	presentation, duration, and spontaneous termination ⁴ , but AF is heterogeneous with regard to
70	clinical presentation, progression, and outcomes.
71	According to the characteristics of AF episodes, the classification subgroups of AF could be
72	first diagnosed as paroxysmal, persistent, and permanent. However, recommendations for AF
73	management do not rely on the temporal AF patterns, except for the restoration of sinus
74	rhythm ^{5,6} . Considering that the pathophysiological mechanism and risk factors for AF are
75	complex, multifactorial, and interacting ^{7,8} , personalized assessment of the pathophysiological
76	process involved in the individual patient by using clinical characteristics, blood biomarkers,
77	and non-invasive substrate determination may improve classification. Further, the guidelines
78	suggested that personalizing management decisions for patients should be based on a
79	structured characterization of AF, including four AF domains-Stroke risk, Symptom severity,
80	Severity of AF burden, and Substrate severity ⁹ . Assessment tools have been established for
81	the first three domains (CHA ₂ DS ₂ -VASC score 10 for Stroke risk, EHRA symptom score 11 for
82	Symptom severity, ICD code for Severity of AF burden). However, there are no clear
83	quantitative criteria for evaluating Substrate severity.

84	Our previous study provided comprehensive information about the associations of risk factor
85	profiles with incident AF ¹² . Based on these risk factors, we aimed to set up a refined
86	classification based on unsupervised and data-driven cluster analyses in UK population that
87	could identify subgroups at different risks of different complications, to help enable
88	personalized intervention. (Supplementary Figure S1).
89	
90	Methods
91	Study population
92	A total of 499,980 adults aged 40-69 years old were recruited from the UK Biobank at
93	baseline ¹³ . We declare that all data are publicly available in the UK Biobank (UKB)
94	repository ¹³ . Among them, 8,293 participants were diagnosed with AF before baseline. After
95	excluding 902 participants for missing data on risk factors, 7,391 participants with AF were
96	included in the main analysis (Supplementary Figure S1), with a follow-up of 86,860
97	person-years. Further, 2,399 participants with newly diagnosed AF within two years after
98	baseline were extracted as a cohort for replication analyses.
99	
100	Ascertainment of incident AF, AF complications and death
101	Incident AF cases (both primary and secondary diagnoses) were identified from the "first
102	occurrence of health outcomes defined by International Statistical Classification of Diseases

- 103 and Related Health Problems, Tenth Revision (ICD-10) (field identification [ID] 131350,
- 104 ICD-10 I48). Similarly, information on acute myocardial infarction (AMI), heart failure (HF),
- 105 and cerebral ischemic stroke (CIS) was obtained by corresponding ICD-10 codes

- 106 (Supplementary Table S1). Information on all-cause and cardiovascular deaths of patients
- 107 was obtained from the death register records.

108

109 **Definition of risk factors**

110 The 2020 European Society of Cardiology (ESC) Guidelines for the diagnosis and 111 management of AF developed in collaboration with the European Association for Cardio-112 Thoracic Surgery summarized 46 risk factors for incident AF, primarily covering 113 demographic factors, health behaviors, clinical comorbidities, cardiometabolic factors and 114 genetic factors³. 115 After an exclusion for unavailable/inadequate data and information integration, we included 116 25 risk factors in the present study (Supplementary Table S2), covering demographic factors 117 (age, sex, race, education, Townsend deprivation index (TDI), air pollution), health behaviors 118 (current smoking, excessive alcohol intake, physical inactivity, excessive physical activity, 119 loneliness, broad depression), clinical comorbidities (hypertension, diabetes/prediabetes, 120 overweight/obesity, low low-density lipoprotein cholesterol (LDL-C), low triglycerides, 121 elevated CRP, sleep apnea, obstructive pulmonary disease (COPD), renal dysfunction, 122 cardiovascular conditions, immune mediated disease, acute illness) and genetic factor (genetic 123 risk score). Detailed information on rationale for risk factor selection and definition of 124 individual risk factors in UK Biobank dataset has been described in our previous study¹². 125

126 Selection of risk factors

127 We used the Least Absolute Shrinkage and Selection Operator (LASSO) regression with 10-

133	subsequent analyses (Supplementary Figure S1 and Table S3).
132	(Supplementary Figures S2 and S3). 23 factors with nonzero coefficient were selected for
131	the largest lambda as possible to obtain the leanest model while ensuring a good-fitting model
130	based on clinical significance. Details were shown in Supplementary Method 1. We chose
129	facilitate easy interpretation, continuous variables were coded as binary categorical variables
128	fold cross-validated ¹⁴ to screen for the principal risk factors for AF, among all 25 factors . To

134

135 Categorization of risk factors

136 Because of the complexity of AF onset and the interaction of risk factors, we used factor 137 analysis to categorize risk factors ¹⁵. Given that four non-modifiable factors (age, sex race, 138 GRS) had a relatively independent role in AF occurrence, and three social-level factors 139 (education, TDI, air pollution) were difficult to change on patients' own, we excluded these 140 seven risk factors in factor analysis. Then, we grouped the remaining 16 risk factors into 6 141 categories (Supplementary Figure 1, Supplementary Method 2). Category 1 was named as 142 metabolic factors, including hypertension, overweight/obesity, and diabetes/prediabetes. 143 Category 2 was named as respiratory factors, including COPD, sleep apnea, current smoking, 144 and elevated CRP. Category 3 was named as cardiovascular factors, including cardiovascular 145 diseases, low levels of LDL-C and triglycerides. Category 4 was named as renal and 146 immunity diseases, including renal dysfunction, immune-mediated disease, and excessive 147 alcohol intake. Category 5 was named as mental health, including loneliness and broad 148 depression. Category 6 was named as acute illness. Finally, we established a score of each 149 category for all participants by summing all risk factors within it, to replace the values

- obtained in the factor analysis considering the convenience of clinical application
 (Supplementary Table S4). A higher score indicated a higher risk of the corresponding risk
 profile.
- 153
- 154 Cluster analysis

155 To focus on risk factors at the individual levels, the 3 social-level risk factors mentioned 156 above were also not to be considered as cluster variables. Sex and race were excluded, as 157 binary variables are not fit for cluster analysis. GRS is not a routinely available variable and 158 was excluded to increase the generalizability of the algorithm. Finally, age and six category 159 scores were chosen for cluster variables, and sex, race, GRS, and 3 social risk factors were 160 included as covariates in the following analyses (Supplementary Figure S1). Cluster 161 variables were standardized to a mean of 0 and standard deviation (SD) of 1 before cluster 162 Consensus clustering analysis was employed to estimate the optimal cluster analysis. 163 number¹⁶. We specified a range of cluster numbers (K=2-12), and iteratively applied the 164 consensus clustering algorithm for each K value. K of 5 was determined as the optimal for the 165 clustering model. More details were shown in Supplementary Method 3. The cluster 166 stability was assessed by Jaccard means, which reflected the consistency of participants in a 167 cluster among 1000 times random resampling.

168

169 Characterization and prognostic analysis

170 Cohen's effect size (d) was calculated to indicate the standardized difference of characters171 between each two clusters. The risk of death and complications was calculated using Cox

172	regression, adjusting sex and race in the basic model, and further adjusting GRS, education,
173	TDI, and air pollution in the additional model. Years of life lost were calculated as the
174	difference in the areas under the survival curves of each cluster ¹⁷⁻¹⁹ . Hazard ratios for death at
175	specified age (interactions between age and cluster were added in Cox regression), prevalence
176	of each AF cluster by age, and population all-cause mortality rates (driven from the Office for
177	National Statistics ²⁰) were used to estimate reductions in life expectancy associated with AF
178	cluster ¹⁷ . In addition, we also used 491,291 participants without AF within two years after
179	baseline as the reference in the supplementary analysis.
180	In addition, we examined the risk of death and complications, and years of life lost for each
181	AF cluster, compared with participants without AF in 2 years follow-up. Associations
182	between clusters and genotypes were calculated with PLINK2.0 ²¹ , after adjustment of age,
183	sex, assessment centers, top 10 principal components of ancestry, and dominance-deviation.
184	To increase the practicability of the clustering algorithm and validate its replicability in
185	newly-onset AF cohorts. First, we established clustering centers for the five clusters in the
186	main analysis. Second, to verify the accuracy of these clustering centers, we re-clustered the
187	participants in the main analysis according to the shortest Pearson distance to the above
188	clustering center. Third, we further classified newly-onset AF patients who developed AF
189	within 2 years of follow-up using established clustering centers. Finally, we assessed the
190	characteristics of these five clusters and the differences in the risk of death and complications
191	among them. Detailed algorithm of clustering AF patients was available on GitHub
192	(https://github.com/JiangLi-1/clustercnp).

194 Results

195 **Distribution of features by clusters**

196	We employed the K-means algorithm to perform clustering on participants with AF at
197	baseline. The Jaccard similarity coefficients in all clusters consistently exceeded 0.76.
198	Demographic, anthropometric, and clinical data for the five clusters are shown in Table 1.
199	The five clusters exhibited distinctive patterns, which were characterized by the standardized
200	means of the cluster variables (Figure 1). Cluster 1, including 587 (7.9%) participants, was
201	characterized by young age and most likely to have acute illnesses, and labelled as acute
202	illness-related AF. Cluster 2, including 1,593 (21.6%), was characterized by poor mental
203	health and moderately likely to have acute illnesses, and labelled as mental health-related AF.
204	Cluster 3 was labelled as cardiovascular disease (CVD)-related AF, including 2,912 (39.4%)
205	participants and characterized by high prevalence of CVD and acute illnesses. Cluster 4 was
206	labelled as immune-and-renal disease related AF, including 1,253 (17.0%) participants and
207	characterized by high prevalence of immune disease and renal dysfunction. The 1,046 (14.2%)
208	participants in cluster 5 (labelled as respiratory-and-metabolic disease-related AF) were
209	characterized by prominent respiratory and metabolic problems. The pairwise comparisons of
210	the clustering variables between clusters were shown in Supplementary Figure S4. Most of
211	the results achieved Bonferroni corrected statistical significance ($P < 0.005$).

212

213

_

Characteristics Cluster 1		Cluster 2	Cluster 3	Cluster 4	Cluster 5
	(n=587)	(n=1,593)	(n=2,912)	(n=1,253)	(n=1,046)
Age	51.68(±5)	60.62(±6.82)	64.07(±3.75)	61.9(±5.91)	64.37(±3.55)
Male sex	422(71.89%)	1000(62.77%)	2130(73.15%)	907(72.39%)	704(67.3%)

Table 1. Basic characteristics for the five clusters

Caucasian race	497(84.67%)	1362(85.5%)	2623(90.08%)	1110(88.59%)	934(89.29%)
Less education	307(52.3%)	987(61.96%)	1755(60.27%)	792(63.21%)	744(71.13%)
Socioeconomic deprivation	125(21.29%)	434(27.31%)	465(15.98%)	224(17.89%)	232(22.22%)
Severe air pollution	144(25%)	412(26.21%)	601(20.93%)	264(21.36%)	233(22.53%)
Metabolic factors					
Hypertension	379(64.57%)	1184(74.33%)	2280(78.3%)	983(78.45%)	887(84.8%)
Overweight/Obesity	448(76.32%)	1224(76.84%)	2195(75.38%)	986(78.69%)	897(85.76%)
Diabetes mellitus/prediabetes	115(19.59%)	545(34.21%)	938(32.21%)	444(35.43%)	482(46.08%)
Respiratory factors					
COPD	24(4.09%)	95(5.96%)	27(0.93%)	63(5.03%)	190(18.16%)
Sleep apnea	14(2.39%)	28(1.76%)	5(0.17%)	26(2.08%)	49(4.68%)
Current smoking	41(6.98%)	118(7.41%)	9(0.31%)	65(5.19%)	135(12.91%)
Elevated CRP	153(26.06%)	401(25.17%)	154(5.29%)	363(28.97%)	883(84.42%)
Cardiovascular factors					
Cardiovascular diseases	180(30.66%)	545(34.21%)	1221(41.93%)	367(29.29%)	422(40.34%)
Low LDL-C	171(29.13%)	609(38.23%)	1223(42%)	475(37.91%)	397(37.95%)
Low triglycerides	118(20.1%)	233(14.63%)	498(17.1%)	208(16.6%)	109(10.42%)
Renal and immunity diseases					
Renal dysfunction	58(9.88%)	213(13.37%)	498(17.1%)	251(20.03%)	288(27.53%)
Immune mediated disease	79(13.46%)	199(12.49%)	372(12.77%)	232(18.52%)	157(15.01%)
Excessive alcohol intake	248(42.25%)	435(27.31%)	1068(36.68%)	534(42.62%)	351(33.56%)
Mental health					
Loneliness	28(4.77%)	1056(66.29%)	42(1.44%)	127(10.14%)	5(0.48%)
Broad depression	12(2.04%)	797(50.03%)	19(0.65%)	88(7.02%)	5(0.48%)
Acute illness					
Acute illness	587(100%)	1274(79.97%)	2912(100%)	171(13.65%)	1046(100%)

214 Data are expressed as mean ± SD or number (percentage). COPD, chronic obstructive

215 pulmonary disease; LDL-C, low-density lipoprotein cholesterol; CRP, C-reactive protein.



216

Figure 1. Cluster features differed among five clusters. All the values of cluster features were centered to a mean value of 0 and an SD of 1. All the negative values were converted to

219 positive values by adding a fixed value to ensure that no cluster feature was below 0.

220

221 Risk of death and major complications among AF clusters

222 During mean follow-up of 11.8 (SD 2.8) years, 1462 (19.8%) deaths were recorded. The risk

223 of all-cause mortality gradually increased from cluster 1 to cluster 5. Compared with

- 224 participants in cluster 1, the risk of death was increased stepwise for those in other clusters
- 225 (cluster 2 to 5: HR, 95%CI 1.85, 1.40-2.43; 1.95, 1.49-2.53; 1.95, 1.47-2.58; 3.21, 2.44-4.23),
- after adjusting sex and race (Figure 2 and Supplementary Table S5-S6). For cardiovascular

death, we found the similar trend as all-cause death (Figure 2 and Supplementary Table S5-

228 **S6**).

229	The association between AF clusters and major complications was shown in Figure 2 and
230	Supplementary Table S7. Compared with participants in cluster 1, those in other clusters had
231	significantly higher risk of AMI, stepwise increased from cluster 4 (1.45, 1.09-1.91), cluster 2
232	(1.75, 1.34-2.29), cluster 3 (1.97, 1.53-2.55), to cluster 5 (2.17, 1.65-2.86). Compared with
233	participants in cluster 1, those in cluster 3 and 5 were at significantly higher risk of HF
234	(cluster 3 and 5: 1.23, 1.03-1.48; 1.67, 1.37-2.03). Additionally, cluster 2, 3 and 5 were at
235	significantly higher risk of CIS (cluster 2, 3 and 5: 1.56, 1.07-2.27; 1.54, 1.08-2.21; 2.00,
236	1.36-2.95), compared with cluster 1. After further adjusting for GRS, education, TDI, and air
237	pollution, the results were similar (Supplementary Table S8).



240	Figure 2. Difference in risk of death and major complications among five clusters. Hazard
241	ratios of all-cause death, cardiovascular death, AMI, HF and CIS for AF cluster 2 to 4,
242	comparing with cluster 1. Cox proportional hazards models were adjusted for sex and race.
243	AMI, acute myocardial infarction, HF, heart failure, CIS, cerebral ischemic stroke.
244	
245	Life expectancy among AF clusters
246	Compared with participants in cluster 1, reduction of life expectancy stepwise increased from
247	cluster 4, cluster 3, cluster 2 and cluster 5 (Figure 3, Supplementary Table S9). For example,
248	at the age of 40 years, we projected a life loss of 32.6 years (95%CI 31.1-34.1) and 30.3

- 249 (29.0-31.9) for participants in cluster 5 and cluster 2, respectively, which was significantly
- 250 higher than those in other two clusters.
- 251



253 Figure 3. Reductions of life expectancy for participants in AF cluster 2 to 4, comparing with 254 cluster 1. The model was adjusted for sex and race. Cluster 2; yellow, cluster 3; brown, 255 cluster 4; red, cluster 5.

256

257 Cluster migration patterns from clinical characteristics to new clusters

- 258 The Sankey diagram shows the patterns of redistribution from classification of clinical
- 259 characteristics to new clusters (Figure 4). Types of AF under the ICD classification had
- 260 different tendencies in new clusters (Supplementary Table S10). Paroxysmal AF preferred to
- be classified in cluster 3 (CVD-related AF) and 4 (immune-and-renal disease related AF)
- 262 (36.1% and 22.3%). However, persistent and chronic AF were prone to be classified into
- 263 cluster 2 (mental health-related AF) and 3 (immune-and-renal disease related AF) (persistent
- AF: 39.7% and 17.2%; chronic AF: 31.3% and 25%). Notably, unspecified AF by clinic
- 265 characters were redistributed into different clusters.





268 Figure 4. Cluster migration pattern from clinical diagnosis to five clusters

269



271	Compared with non-AF participants, all clusters had significantly higher risk of death and								
272	major comp	major complications (Supplementary Table S11). Among three complications, the risk of HF							
273	was highest	vas highest for all clusters. Compared with non-AF participants, individuals in cluster 2 and 5							
274	at age 40 y	at age 40 years were projected to die about 30 years earlier than those without AF in 2 years							
275	follow-up,	and the numb	ers in cluster	1, 3 and 4 were	e about 10 yea	rs (Figure 5 a	and		
276	Supplementary Table S12). In addition, we analysed the association of genetic loci risk for								
277	AF ²² with each AF cluster, compared with non-AF participants (Supplementary Table S13).								
278	The association of three genetic variants (rs11264280, rs2129977, rs2359171) with at least								
279	one cluster reached the genome-wide level of significance ($P < 5*10^{-8}$) and differed among								
280	five clusters (Table 2).								
281									
282	Table 2. G	enetic association	ons of risk vari	ants for AF read	ching significanc	ce of genome-w	vide		
283	association	study with at lea	st one cluster						
SNP	Nearest GENE	Cluster 1 BETA (P)	Cluster 2 BETA (P)	Cluster 3 BETA (P)	Cluster 4 BETA (P)	Cluster 5 BETA (P)	Pinterac		

SINF	GENE	BETA (P)	r interaction				
rs11264280	KCNN3	0.0001797	0.0007819	0.00146185	0.0005184	0.0004884	5.57E-26
		p=0.044	p=8.81E-08	p=8.80E-14	p=6.43E-05	p=4.03E-05	
rs2129977	PITX2	0.0010089	0.0020606	0.00356029	0.0018736	0.0015124	1.52E-86
		p=4.89E-15	p=1.85E-22	p=2.56E-36	p=1.58E-23	p=1.13E-18	
rs2359171	ZFHX3	0.0003318	0.0011624	0.00173702	0.000418	0.0011117	2.81E-17
		p=0.028	p=2.71E-06	p=1.73E-07	p=0.058	p=3.16E-08	

The generalized linear model was used to estimate the associations of each risk variant for AF with the five clusters, after adjustment of age, sex, assessment centres, top 10 principal components of ancestry, and dominance-deviation. The P_{interaction} was based on the assessment of the significance in the differences of associations between the specific risk variant and the five clusters.

289

290 Practicality of the clustering algorithm and replication in a new cohort

291 The characteristics were similar between original clusters in main analysis and re-clusters that

292 used the shortest Pearson distance from the clustering center (Supplementary Figure S6 and

293	Table S14). Then, we clustered participants with newly-onset AF in 2 years after baseline and
294	obtained similar five clusters to the original clusters (Supplementary Figure S7). In addition,
295	we found that the difference in disease risks were similar to those in main analyses
296	(Supplementary Table S15).

$^{\circ}$	n	7
4	Э	1

298 Discussion

299 Our findings provided a new clustering of AF patients, which might help to identify patients 300 at high risk for AF related complications and guide the choice of treatment. We found that AF 301 patients could be classified into five clusters of risk profiles. Cluster 1 (acute illness-related 302 AF) was characterized by young age and acute illnesses, cluster 2 (mental health-related AF) 303 by poor mental health, cluster 3 (CVD-related AF) by a high prevalence of cardiovascular 304 diseases and acute illnesses, cluster 4 (immune-and-renal disease related AF) by high 305 prevalence of immune disease and renal dysfunction and cluster 5 (respiratory-and-metabolic 306 disease-related AF) by prominent respiratory and metabolic problems. In addition to 307 population characteristics, the risk of death and important complications, as well as loss of 308 life expectancy, differed between clusters. The underlying genetic background behind the five 309 clusters also varied. Furthermore, we applied this cluster algorithm based on the existing 310 model to classify AF patients in clinical setting.

Among five clusters, cluster 5 (respiratory- and metabolic disease-related AF) represented the most severe type of AF, and this cohort deserved more attention, intensified treatment, and long-term follow-up to reduce the risk of AF complications. Patients in cluster 5 had a significantly higher risk of death and reduction of life expectancy than the other clusters,

315	which underscored the fact that metabolic and respiratory co-morbidities should be prioritized
316	in managing AF patients. On one hand, COPD and apnea, as two respiratory factors strongly
317	associated with AF, both contribute to hypoxemia. Patients with underlying cardiac disease
318	including AF have a higher risk of death and AMI under the condition of prolonged
319	hypoxemia ²³ . On the other hand, many metabolic risk factors contribute to the development
320	of AF ²⁴ , leaded by overweight/obesity ¹² . We identified that ZFHX3 (rs2359171) was only
321	associated with cluster 5. ZFHX3 was reported in genome-wide association analyses of both
322	BMI and AF ^{22,25} , which may be the underlying genetic shared between overweight/obesity
323	and AF. Furthermore, patients in cluster 5 also had highest risk of HF. Anticoagulation to
324	prevent stroke has already been integrated part of Atrial fibrillation Better Care (ABC)
325	holistic pathway ³ . Primary prevention of HF in AF seems to be less considered. We suppose
326	that in patients with respiratory- and metabolic disease-related AF, use of sodium/glucose
327	cotransporter 2 inhibitor might improve HF outcome, which warrants further trials.
328	Notably, we found that individuals in cluster 2 (mental health-related AF) lost the second
329	most life expectancy, out of the five clusters. Cluster 2 had a unique profile, showing the
330	highest prevalence of psychological problems (loneliness and depression) among five clusters.
331	Although psychological problems are not included in the guidelines as risk factors for AF, the
332	role of mental health especially depression in the pathogenesis of AF has been acknowledged,
333	^{26,27} . Our results provided further evidence that KCNN3 (rs2129977) was associated with
334	mental health-related AF. KCNN3, performing the function of promoting action potential
335	repolarization ²⁸ , was proved to be associated with the development of AF 22,29 and influenced
336	by depressive illness ³⁰ , which indicates a potential mechanism by which mood influences the

337 onset of AF. Mental health deserves more attention in the prevention and control of AF.

- 338 Further, the prevention of CIS should be attached more importance in those patients,339 considering the second higher risk of CIS among the five clusters.
- 340 We found that cluster 1 (acute illness-related AF) was the cluster with the best prognosis 341 among all clusters. This group clusters 45% people younger than 55 years old in this study, 342 suggesting the importance of acute illness on development of AF in young and middle-aged 343 patients. Further, acute illness (major operation, pneumonia) and immune-related diseases are both responsible for inflammation and oxidative stress, and eventually induces AF ³¹⁻³³. 344 345 Therefore, we believe it's worthwhile to control immune response and rationally use 346 antioxidants in these patients. Of note, an increasing number of randomized controlled trials 347 have demonstrated the significant effect of antioxidants and control of the immune response 348 in the prevention of AF ^{34,35}.
- Traditional AF classification is based on presentation, duration and spontaneous termination of AF episodes. However, in clinical practice, a large proportion of patients presented insidious symptoms³⁶, which makes it difficult to be classified by AF episodes. In contrast, the information needed for our classification method is more easily collected and patients with unspecified AF coded by ICD were further classified into specific clusters.
- In the present study, we provided a feasible score of 1 to 5 for patients, classified into clusters 1 to 5 in order, in terms of clinical assessment in Substrate severity of AF. Due to the underlying heterogeneity of the influence of different risk factors on AF, it is not advisable to simply list the number of risk factors in a patient as a risk score. We took the diversity of AF comorbidities and the heterogeneity of their prognostic impact into account. By integrating

359 much available risk factors as possible into the clustering analysis, our classification method

360 can refine AF subtyping and guide more accurate risk stratification.

361	Our study has several limitations. First, we cannot claim that this is the best clustering for AF
362	although we made full use of risk factors from previous evidence. Moreover, whether there is
363	stability of clustering for patients with distant centers of clustering needs to be confirmed in
364	future prospective cohorts. Second, the main applicable ethnicity for our clustering algorithm
365	is European and the stability of this algorithm should be further validated in other cohorts.
366	Third, the generalizability of our findings was limited to population older than 40 because of
367	the composition of the participants in UKB. Further research is warranted in the population
368	presenting early-onset AF.

369

370 Clinical implication

371 The identification of clusters based on demographic, anthropometric, and clinical 372 characteristics allows for a more personalized approach to patient care and interventions 373 based on specific risk profiles. The recognition of different AF clusters ("acute illness-related 374 AF", "mental health-related AF", "CVD-related AF", "immune-and-renal disease related AF", 375 and "respiratory-and-metabolic disease-related AF") can help to guide physicians to 376 understand the diverse pathophysiological mechanisms underlying AF. Strategies of therapy 377 targeting the specific characteristics of each cluster may lead to more effective prevention of 378 complications and outcome improvement. Furthermore, the observation of risk of death and 379 major complications in patients within each AF cluster underscores the importance of risk 380 stratification and long-term monitoring. Patients in clusters with high-risk profiles, such as

- 381 "respiratory- and metabolic disease-related AF", may require the most positive management
- 382 to mitigate adverse outcomes.
- 383

384	Conc	lusion
004	COLLE	usivii

385 Different AF subtypes exhibited unique features and difference in mortality and 386 complications. By integrating health habits, demographic characteristics, biochemical 387 indicators, and disease diagnostic information, we provide clinically valuable clustering in 388 patients with AF, gives the opportunity to benefit from precision interventions.

389

390 Ethnic declaration

391 The UKB received ethical approval from the UK National Health Service, National Research

392 Ethics Service North West, the National Information Governance Board for Health and Social

393 Care in England and Wales, and the Community Health Index Advisory Group in Scotland.

394 All procedures followed were in accordance with the ethical standards of the responsible

395 committee on human experimentation (institutional and national) and with the Helsinki

396 Declaration of 1975, as revised in 2008. All participants provided written informed consent.

397 This study was approved by the UK Biobank (application number 77740).

398

399 Data availability statement

400 The data underlying this article are available in UK Biobank, at <u>https://www.ukbiobank.ac.uk/</u>.

401 The datasets were derived from sources in the public domain:

402 https://ukbiobank.dnanexus.com/.

403

404 Code availability

- 405 The R "clustercnp" package for clustering AF patients is available on
- 406 https://github.com/JiangLi-1/clustercnp.
- 407

408 Acknowledgements

- This research has been conducted using the UK Biobank Resource under Application Number
- 410 77740. This study was supported by Shanghai Municipal Health Commission (2022XD017),
- 411 Clinical Research Plan of SHDC (SHDC2020CR4006), National Natural Science Foundation
- 412 of China (82170870), Shanghai Municipal Human Resources and Social Security Bureau
- 413 (2020074), Innovative Research Team of High-level Local Universities in Shanghai
- 414 (SHSMU-ZDCX20212501). The funders played no role in the design or conduction of the
- 415 study; in the collection, management, analysis, or interpretation of the data; or in the
- 416 preparation, review, or approval of the article.
- 417

418 Declarations

- 419 No potential conflicts of interest relevant to this article were reported.
- 420

421 Author contributions

- 422 N.W. and Y.L. concepted this paper. Y.Y., J.L. and N.W. wrote the manuscript, researched data,
- 423 and reviewed/edited the manuscript. N.W. and X.T. reviewed/edited the manuscript. J.L., Y.S.,
- 424 H.Z., Y.W. and B.W. contributed to data acquisition and revised it critically for important

425	intellectual	content.	All	authors	approved	the	final	manuscript.
-----	--------------	----------	-----	---------	----------	-----	-------	-------------

426 Reference

427	1. Benjamin EJ, Muntner P, Alonso A, et al. Heart Disease and Stroke Statistics-2019								
428	Update: A Report From the American Heart Association. <i>Circulation</i> 2019; 139 :e56-e528. doi:								
429	10.1161/cir.000000000000659								
430	2. Roth GA, Mensah GA, Johnson CO, et al. Global Burden of Cardiovascular Diseases								
431	and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol								
432	2020; 76 :2982-3021. doi: 10.1016/j.jacc.2020.11.010								
433	3. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and								
434	management of atrial fibrillation developed in collaboration with the European Association for								
435	Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of								
436	atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special								
437	contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J								
438	2021; 42 :373-498. doi: 10.1093/eurheartj/ehaa612								
439	4. Kirchhof P, Benussi S, Kotecha D, et al. 2016 ESC Guidelines for the management of								
440	atrial fibrillation developed in collaboration with EACTS. Eur Heart J 2016; 37 :2893-2962. doi:								
441	10.1093/eurheartj/ehw210								
442	5. January CT, Wann LS, Calkins H, et al. 2019 AHA/ACC/HRS Focused Update of the								
443	2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: A								
444	Report of the American College of Cardiology/American Heart Association Task Force on								
445	Clinical Practice Guidelines and the Heart Rhythm Society in Collaboration With the Society								
446	of Thoracic Surgeons. <i>Circulation</i> 2019; 140 :e125-e151. doi:								
447	10.1161/CIR.000000000000665								

448	6. Group NCAFGW, Brieger D, Amerena J, et al. National Heart Foundation of Australia							
449	and the Cardiac Society of Australia and New Zealand: Australian Clinical Guidelines for the							
450	Diagnosis and Management of Atrial Fibrillation 2018. Heart Lung Circ 2018;27:1209-1266.							
451	doi: 10.1016/j.hlc.2018.06.1043							
452	7. Lau DH, Schotten U, Mahajan R, et al. Novel mechanisms in the pathogenesis of atrial							
453	fibrillation: practical applications. <i>Eur Heart J</i> 2016; 37 :1573-1581. doi:							
454	10.1093/eurheartj/ehv375							
455	8. Staerk L, Sherer JA, Ko D, Benjamin EJ, Helm RH. Atrial Fibrillation: Epidemiology,							
456	Pathophysiology, and Clinical Outcomes. <i>Circ Res</i> 2017; 120 :1501-1517. doi:							
457	10.1161/circresaha.117.309732							
458	9. Potpara TS, Lip GYH, Blomstrom-Lundqvist C, et al. The 4S-AF Scheme (Stroke Risk;							
459	Symptoms; Severity of Burden; Substrate): A Novel Approach to In-Depth Characterization							
460	(Rather than Classification) of Atrial Fibrillation. <i>Thromb Haemost</i> 2021; 121 :270-278. doi:							
461	10.1055/s-0040-1716408							
462	10. Kirchhof P, Benussi S, Kotecha D, et al. 2016 ESC Guidelines for the management of							
463	atrial fibrillation developed in collaboration with EACTS. <i>Europace</i> 2016; 18 :1609-1678. doi:							
464	10.1093/europace/euw295							
465	11. Björkenheim A, Brandes A, Magnuson A, et al. Assessment of Atrial Fibrillation-Specific							
466	Symptoms Before and 2 Years After Atrial Fibrillation Ablation: Do Patients and Physicians							
467	Differ in Their Perception of Symptom Relief? JACC Clin Electrophysiol 2017;3:1168-1176.							
468	doi: 10.1016/j.jacep.2017.04.003							
469	12. Wang N, Yu Y, Sun Y, et al. Acquired risk factors and incident atrial fibrillation according							

- 470 to age and genetic predisposition. *Eur Heart J* 2023. doi: 10.1093/eurheartj/ehad615
- 471 13. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for
- 472 identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*
- 473 2015; 12:e1001779. doi: 10.1371/journal.pmed.1001779
- 474 14. Tibshirani R. Regression Shrinkage and Selection via The Lasso: A Retrospective.
- 475 Journal of the Royal Statistical Society Series B: Statistical Methodology 2011;73:273-282.
- 476 doi: 10.1111/j.1467-9868.2011.00771.x
- 477 15. Gaskin CJ, Happell B. On exploratory factor analysis: a review of recent evidence, an
- 478 assessment of current practice, and recommendations for future use. Int J Nurs Stud
- 479 2014;51:511-521. doi: 10.1016/j.ijnurstu.2013.10.005
- 480 16. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based
- 481 Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine
- 482 Learning 2003;52:91-118. doi: 10.1023/A:1023949509487
- 483 17. Life expectancy associated with different ages at diagnosis of type 2 diabetes in high-
- 484 income countries: 23 million person-years of observation. *Lancet Diabetes Endocrinol* 2023.
- 485 doi: 10.1016/s2213-8587(23)00223-1
- 486 18. Xu C, Zhang P, Cao Z. Cardiovascular health and healthy longevity in people with and
- 487 without cardiometabolic disease: A prospective cohort study. EClinicalMedicine
- 488 2022;45:101329. doi: 10.1016/j.eclinm.2022.101329
- 489 19. Sun Y, Yu Y, Zhang K, et al. Association between Life's Essential 8 score and risk of
- 490 premature mortality in people with and without type 2 diabetes: A prospective cohort study.
- 491 Diabetes Metab Res Rev 2023; **39**:e3636. doi: 10.1002/dmrr.3636

492	20.	Census.	Single-year	life	tables,	UK:	1980	to	2020.
-----	-----	---------	-------------	------	---------	-----	------	----	-------

- 493 https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpect
- 494 ancies/datasets/singleyearlifetablesuk1980to2018/singleyearlifetablesuk
- 495 21. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge
- 496 of larger and richer datasets. *Gigascience* 2015;4:7. doi: 10.1186/s13742-015-0047-8
- 497 22. Roselli C, Chaffin MD, Weng LC, et al. Multi-ethnic genome-wide association study for
- 498 atrial fibrillation. Nat Genet 2018;50:1225-1233. doi: 10.1038/s41588-018-0133-9
- 499 23. Yeghiazarians Y, Jneid H, Tietjens JR, et al. Obstructive Sleep Apnea and
- 500 Cardiovascular Disease: A Scientific Statement From the American Heart Association.
- 501 *Circulation* 2021;**144**:e56-e67. doi: 10.1161/cir.00000000000988
- 502 24. Braunwald E. Diabetocardiology: a new subspecialty? European Heart Journal
- 503 2023:ehad541. doi: 10.1093/eurheartj/ehad541
- 504 25. Turcot V, Lu Y, Highland HM, et al. Protein-altering variants associated with body mass
- 505 index implicate pathways that control energy intake and expenditure in obesity. Nat Genet
- 506 2018;50:26-41. doi: 10.1038/s41588-017-0011-x
- 507 26. Sadlonova M, Wasser K, Nagel J, et al. Health-related quality of life, anxiety and
- 508 depression up to 12 months post-stroke: Influence of sex, age, stroke severity and atrial
- 509 fibrillation A longitudinal subanalysis of the Find-AF(RANDOMISED) trial. J Psychosom Res
- 510 2021; **142**: 110353. doi: 10.1016/j.jpsychores.2020.110353
- 511 27. Feng T, Malmo V, Laugsand LE, et al. Symptoms of anxiety and depression and risk of
- 512 atrial fibrillation-The HUNT study. Int J Cardiol 2020;306:95-100. doi:
- 513 10.1016/j.ijcard.2019.11.107

- 514 28. Zhang X-D, Lieu DK, Chiamvimonvat N. Small-conductance Ca2+-activated K+ channels
- 515 and cardiac arrhythmias. Heart Rhythm 2015;12:1845-1851. doi:
- 516 https://doi.org/10.1016/j.hrthm.2015.04.046
- 517 29. Rahm AK, Wieder T, Gramlich D, et al. HDAC2-dependent remodeling of K(Ca)2.2
- 518 (KCNN2) and K(Ca)2.3 (KCNN3) K(+) channels in atrial fibrillation with concomitant heart
- 519 failure. *Life Sci* 2021;**266**:118892. doi: 10.1016/j.lfs.2020.118892
- 520 30. Sargin D, Oliver DK, Lambe EK. Chronic social isolation reduces 5-HT neuronal activity
- 521 via upregulated SK3 calcium-activated potassium channels. *Elife* 2016;5. doi:
- 522 10.7554/eLife.21416
- 523 31. Zakkar M, Ascione R, James AF, Angelini GD, Suleiman MS. Inflammation, oxidative
- 524 stress and postoperative atrial fibrillation in cardiac surgery. *Pharmacol Ther* 2015; **154**:13-20.
- 525 doi: 10.1016/j.pharmthera.2015.06.009
- 526 32. Sagris M, Vardas EP, Theofilis P, et al. Atrial Fibrillation: Pathogenesis, Predisposing
- 527 Factors, and Genetics. Int J Mol Sci 2021;23. doi: 10.3390/ijms23010006
- 528 33. Yang KC, Dudley SC, Jr. Oxidative stress and atrial fibrillation: finding a missing piece to
- 529 the puzzle. *Circulation* 2013;**128**:1724-1726. doi: 10.1161/circulationaha.113.005837
- 530 34. Rodrigo R, Korantzopoulos P, Cereceda M, et al. A randomized controlled trial to prevent
- 531 post-operative atrial fibrillation by antioxidant reinforcement. J Am Coll Cardiol 2013;62:1457-
- 532 1465. doi: 10.1016/j.jacc.2013.07.014
- 533 35. Zuo H, Nygård O, Ueland PM, et al. Association of plasma neopterin with risk of an
- 534 inpatient hospital diagnosis of atrial fibrillation: results from two prospective cohort studies. J
- 535 Intern Med 2018;283:578-587. doi: 10.1111/joim.12748

- 536 36. Schnabel RB, Marinelli EA, Arbelo E, et al. Early diagnosis and better rhythm
- 537 management to improve outcomes in patients with atrial fibrillation: the 8th AFNET/EHRA
- 538 consensus conference. *Europace* 2023;25:6-27. doi: 10.1093/europace/euac062

539









Age



Acute illness

Mental health

Immune and renal disease



Risk for all-cause death (HR [95%CI])



Risk for AMI (HR [95%CI])

Risk for Cardiovascular death (HR [95%CI])

D

Risk for HF (HR [95%CI])

1.0 (REF)

- 1.57 (1.13,2.19)
- 1.83 (1.34,2.50)
- 1.88 (1.35,2.61)
- 2.90 (2.09,4.03)

Risk for CIS (HR [95%CI])



Age at risk (years)



Persistent AF Chronic AF

Unspecified AF

