

1 ClustALL: A robust clustering strategy for stratification of patients with acutely 2 decompensated cirrhosis

3
4 Sara Palomino-Echeverria, spalomie@navarra.es, Spain¹
5 Estefania Huergo, ehuergoi@navarra.es, Spain¹
6 Asier Ortega-Legarreta, aortegal@navarra.es, Spain¹
7 Eva M. Uson, eva.uson@efclif.com, Spain²
8 Ferran Aguilar, ferran.aguilar@efclif.com, Spain²
9 Carlos de la Pena, carlos.delapena@efclif.com, Spain²
10 Cristina Lopez-Vicario, clopezv@recerca.clinic.cat, Spain³
11 Carlo Alessandria, carloalessandria@libero.it, Italy⁴
12 Wim Laleman, wim.laleman@uzleuven.be, Belgium⁵
13 Alberto Farias Queiroz, alberto.farias@fm.usp.br, Brazil⁶
14 Richard Moreau, richard.moreau@inserm.fr, France^{2,7,8,9}
15 Javier Fernandez, javier.fernandez@efclif.com, Spain²
16 Vicente Arroyo, vicente.arroyo@efclif.com, Spain²
17 Paolo Caraceni, paolo.caraceni@unibo.it, Italy¹⁰
18 Vincenzo Lagani, vincenzo.lagani@gmail.com, Georgia^{11,12,13}
19 Cristina Sanchez, cristina.sanchez@efclif.com, Spain²
20 Joan Claria, jclaria@clinic.cat, Spain³
21 Jesper Tegner, jesper.tegner@kaust.edu.sa, Saudi Arabia¹¹
22 Jonel Trebicka, jonel.trebicka@ukmuenster.de, Germany^{2,14}
23 Narsis Kiani, narsis.kiani@ki.se, Sweden¹⁵
24 Nuria Planell, nplanellpic@unav.es, Spain^{1,16*}
25 Pierre-Emmanuel Rautou, pierre-emmanuel.rautou@inserm.fr, France^{7,17*}
26 David Gomez-Cabrero, david.gomez.cabrero@navarra.es, Spain^{1,11*}

27
28 **PREDICT investigators, ACLARA investigators, DECISION investigators.**

29 **Affiliations:**

30
31 1: Unit of Translational Bioinformatics, Navarrabiomed - Fundación Miguel Servet, Pamplona, Spain
32 2: European Foundation for the Study of Chronic Liver Failure, Barcelona, Spain
33 3: Biochemistry and Molecular Genetics Service, Hospital Clínic-IDIBAPS, Barcelona, Spain; European Foundation for the
34 Study of Chronic Liver Failure (EF Clif) and Grifols Chair, Barcelona, Spain; CIBERehd, Barcelona, Spain; Department of
35 Biomedical Sciences, University of Barcelona, Barcelona, Spain
36 4: A.O.U. Città della Salute e della Scienza Torino, Torino, Italy
37 5: University of Leuven, Leuven, Belgium
38 6: Department of Gastroenterology, Hospital das Clínicas, University of São Paulo School of Medicine, Brazil
39 7: Université Paris-Cité, Inserm, Centre de recherche sur l'inflammation, UMR, Paris, France.
40 8: Assistance Publique-Hôpitaux de Paris (AP-HP), France
41 9: Hôpital Beaujon, Service d'Hépatologie, Clichy, France
42 10: University of Bologna, Department of Medical and Surgical Science, Bologna, Italy
43 11: Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology
44 (KAUST), Thuwal, Saudi Arabia.
45 12: SDAIA-KAUST Center of Excellence in Data Science and Artificial Intelligence, Thuwal, Saudi Arabia
46 13: Institute of Chemical Biology, Ilia State University, Tbilisi 0162, Georgia
47 14: Department of internal medicine B, University of Münster, Münster, Germany
48 15: Algorithmic Dynamics lab, Karolinska Institutet, Solna, Sweden
49 16: Universidad de Navarra, CIMA, Computational Biology Program, Instituto de Investigación Sanitaria de Navarra
50 (IdiSNA), 31008 Navarra, Spain.
51 17: Service d'Hépatologie, AP-HP, Hôpital Beaujon, DMU DIGEST, Centre de Référence des Maladies Vasculaires du Foie,
52 FILFOIE, ERN RARE-LIVER, Clichy, France.
53 * Contributed equally.
54
55

56 Abstract

57

58 Patient heterogeneity represents a significant challenge for both individual patient management and
59 clinical trial design, especially in the context of complex diseases. Most existing clinical classifications
60 are based on scores built to predict patients' outcomes. These classical methods may thus miss
61 features that contribute to heterogeneity without necessarily translating into prognostic
62 implications.

63 To address patient heterogeneity at hospital admission, we developed ClustALL, a computational
64 pipeline designed to handle common clinical data challenges such as mixed data types, missing
65 values, and collinearity. ClustALL also facilitates the unsupervised identification of multiple and
66 robust stratifications. We applied ClustALL to a prospective European multicentre cohort of patients
67 with acutely decompensated cirrhosis (AD) (n=766), a highly heterogeneous disease. ClustALL
68 identified five robust stratifications for patients with AD, using only data at hospital admission. All
69 stratifications included markers of impaired liver function and number of organ dysfunction or
70 failure, and most included precipitating events. When focusing on one of these stratifications,
71 patients were categorized into three clusters characterized by typical clinical features but also having
72 a prognostic value. Re-assessment of patient stratification during follow-up delineated patients'
73 outcomes, with further improvement of the prognostic value of the stratification. We validated
74 these findings in an independent prospective multicentre cohort of patients from Latin America
75 (n=580).

76 In conclusion, this study developed ClustALL, a novel and robust stratification method capable of
77 addressing challenges tied to intricate clinical data and applicable to complex diseases. By applying
78 ClustALL to patients with AD, we identified three patient clusters, offering insights that could guide
79 future clinical trial design.

80

81

82 Introduction

83 Heterogeneity is a prevalent phenomenon observed in numerous diseases, including various types of
84 cancer (1), autoimmune conditions like multiple sclerosis (2), and diabetes (3)). This becomes
85 especially critical in diseases where environmental and lifestyle factors play a significant role. Acutely
86 decompensated cirrhosis, which refers to the rapid development of overt ascites, overt hepatic
87 encephalopathy, variceal haemorrhage, or any combination of these disorders, which often leads to
88 nonelective admission to the hospital of patients who were previously stable (4), exemplifies
89 significant inter-individual variability. It encompasses a range of causes of cirrhosis, comorbidities,
90 precipitating events, clinical presentations, and outcomes (4). This clinical heterogeneity poses a
91 considerable challenge as it likely accounts for the diverse responses to treatment and outcomes
92 observed in these patients (5). Therefore, we reasoned that analysing a large population of patients
93 with acutely decompensated cirrhosis should allow us to develop stratification tools.

94
95 A major tool for the characterization of patient heterogeneity is the identification of patient
96 subtypes, also defined as patient stratification. Importantly, the World Health Organization has
97 acknowledged patient stratification as a valuable approach for enhancing population health
98 management and providing better-tailored services (6). In conceptual terms, patient stratification
99 can be described as the process of grouping or clustering patients based on specific characteristics or
100 patterns without relying on labelled data or information about future outcomes (7). Therefore,
101 contrary to scores developed using classical statistical approaches based on the clinical course,
102 stratification can capture features explaining patients' heterogeneity independently of their
103 association with patient outcomes.

104
105 Numerous attempts have been made to identify subgroups within clinical datasets (7–9). However,
106 the lack of a universally applicable approach poses a significant challenge in the field of clustering
107 analysis. Although there have been advancements beyond the classical k-means and hierarchical
108 clustering methods, no general framework still allows the organization and classification of clustering
109 methodologies in the clinical setting (10). Instead, many ad-hoc applications have been developed for
110 specific scenarios, but their generalizability is often limited. While there is no global classification,
111 these applications can be grouped based on specific characteristics such as managing missing values,
112 collinearity, or mixed data (9). For instance, when handling missing data, some methods exclude
113 samples from the analysis, potentially resulting in a loss of statistical power, while others rely on a
114 single imputation, overlooking the potential bias that can be introduced (11). Highly correlated
115 variables represent a challenge. Some methods exclude them, while others employ dimensionality
116 reduction techniques such as Principal Component (PC) reduction to capture underlying lower-
117 dimensional data patterns (12,13). However, both decisions may affect the outcome of the
118 clustering, as sensitivity analyses are rarely conducted. Moreover, indiscriminate feature selection
119 can inadvertently remove informative features along with noisy ones, potentially biasing the results
120 (14). Furthermore, most clustering methodologies assume the existence of a single stratification,
121 disregarding the possibility of having none or multiple valid alternatives for subgrouping the
122 population (15). Interestingly, trace-based clustering methodologies have recently emerged to aid in
123 the interpretation and validation of the identified subgroups, often requiring domain knowledge and
124 expert input (16).

125

126 Additionally, the evaluation of clustering outcomes is an open problem that is based on the quality of
127 the produced clusters. In the case of unsupervised clustering, where no preliminary classification
128 exists, evaluations are typically referenced against theoretical benchmarks. For instance, when
129 addressing the optimal number of clusters, various theoretical quality metrics are available such as
130 the clustering coefficient (17) or the silhouette index (18) among many others. Importantly, while
131 there is no universal methodology that excels across all scenarios for all data sets, as dictated by the
132 "no free lunch" theorem (19), there exist strategies that yield high-quality results (20–22). Another
133 essential measure—referred to as robustness—lacks a precise definition. Robustness, in general
134 terms, signifies the capacity of a system to withstand changes (23). In our context, we investigate
135 whether a clustering remains stable when subjected to perturbations. In this work, we considered
136 two types of perturbations: those derived from changes in the population and those arising from
137 changes in the algorithm's parameters. In the case of population-based perturbations, we quantify
138 how a given clustering is influenced by variations in the underlying population. Bootstrapping is one
139 approach to address this scenario (24). In the case of parameter-based perturbations, we assess the
140 impact of parameter adjustments in the clustering algorithm on the identified clustering (25).
141 Consider a scenario where a parameter "x" defines our clustering strategy. How different is the
142 resulting clustering when using "x=1" versus "x=1.1"? Here, robustness translates to clusterings that
143 maintain stability even when parameter values shift. For the reader's clarity, we will name the two
144 different robustness criteria: **population-based robustness** and **parameter-based robustness**.

145

146 Importantly, there is currently no methodology capable of addressing all the aforementioned
147 scenarios while ensuring both definitions of robustness. To address these challenges
148 comprehensively, we developed ClustALL, a novel framework that robustly identifies patient
149 subgroups by addressing all the previously mentioned challenges and limitations of existing
150 methodologies, and applied ClustALL -as a proof-of-concept- in two large prospective cohorts of
151 patients non-electively admitted to the hospital for acutely decompensated cirrhosis.

152

153 In this study, ClustALL was applied to a large prospective cohort of patients non-electively admitted
154 to the hospital for acutely decompensated cirrhosis. The resulting stratifications were thoroughly
155 characterized, aiming to identify any particular stratification of special interest in the clinical context
156 showing prognostic value. We then validated the reproducibility of this stratification using a separate
157 prospective cohort of patients. One further aim of the study was to demonstrate the usability of
158 stratification over the disease course, with prognostic value.

159

160 Results

161 ClustALL, a robust data-driven framework for patient stratification in complex 162 diseases

163

164 We developed a specialized stratification framework, referred to as ClustALL, specifically designed to
165 accurately identify all potential alternatives for stratifying a population using clinical multimodal data
166 at hospital admission as input. The ClustALL methodology consists of three main steps illustrated in
167 **Figure 1** and detailed in the Methods section: (1) **Data Complexity Reduction** (depicted in the Green
168 Panel of **Fig.1**) aims to simplify the original dataset by mitigating the impact of redundant
169 information (highly correlated variables). As a result, we obtain a set of embeddings, each one
170 derived from different groupings of clinical variables. (2) **Stratification Process** (depicted in the
171 Purple Panel of **Fig.1**), where, for each embedding, multiple stratification analyses are performed
172 using different combinations of distance metrics and clustering methodologies. From each
173 combination, denoted as "*embedding + distance metric + clustering method*", a stratification is
174 derived. (3) **Consensus-based Stratifications** step (depicted in the Red Panel of **Fig.1**) aims to identify
175 robust stratifications that, in addition, exhibit minimal variation when combination parameters
176 ("*embedding + distance metric + clustering method*") are *slightly* modified. ClustALL performs a
177 **population-based robustness** analysis for each stratification using bootstrapping. This analysis
178 ensures that combinations associated with non-robust stratifications are excluded. The resulting
179 stratifications are then compared using the Jaccard distance. As a result, a heatmap is generated to
180 visually identify groups of representative stratifications (green squared lines). The selection of
181 representative stratifications enables the preservation of those stratifications that demonstrate
182 **parameter-based robustness**: consistency even when various parameters, like distance metrics or
183 clustering methods, are altered. For each group of stratifications, the centroid is selected as the final
184 stratification (green squares).

185

186 Combining **these three steps allows ClustALL to identify none, one, or multiple robust**
187 **stratifications in a given population** of patients with complex diseases. Importantly, a specific
188 implementation of ClustALL is designed to **effectively handle datasets with missing data effectively**,
189 ensuring that incomplete information does not hinder the stratification process.

190

191 ClustALL uncovers stratification in a cohort of patients with acutely 192 decompensated cirrhosis: a proof-of-concept

193 Study population.

194 The ClustALL approach was applied to a subset of individuals from the European PREDICT cohort
195 (26), which included 766 patients with acute decompensation of cirrhosis and 74 clinical features
196 collected at hospital admission, with less than 30% missing values. Complete information on patient
197 characteristics and short-term outcomes, including acute-on-chronic liver failure (ACLF), liver
198 transplant, and death, can be found in **Supplemental Table 1**.

199

200 ClustALL identified five different alternatives to stratify the population.

201 The ClustALL workflow was utilized to discover potential new sub-phenotypes of patients with acute
202 decompensation of cirrhosis within the PREDICT cohort upon hospital admission (**Fig.2**). To handle
203 missing values in the dataset, we employed the ClustALL framework, which incorporates imputations
204 using 1,000 iterations, as described in the Methods section. The Data Complexity Reduction Step
205 resulted in 72 embeddings (**Fig.2.1**). The Stratification Process generated 288 stratifications based on
206 the different combinations of "*embedding + distance metric + clustering method*" (**Fig.2.2**). Among
207 these, 144 population-based robust stratifications were identified through the Consensus-based
208 Stratifications step, resulting in five groups of parameter-based representative stratifications. The
209 centroid was selected for each group of stratifications, (**Fig.2.3**).

210

211 ClustALL provides better resolution than classical clustering tools

212 We conducted an analysis to assess the added value of ClustALL when compared with classical
213 clustering methodologies such as k-means or hierarchical clustering. Regarding the classical
214 methodologies, our findings revealed that when using correlation as a distance metric, 90% of
215 patients were consistently assigned to a single cluster, regardless of the number of clusters
216 considered; when Gower distance was utilized, the distribution of patients across clusters presented
217 a more balanced distribution (**Table S2**). Notably, the population-based robustness of the
218 stratifications generated by ClustALL was significantly higher (p-value <0.01) compared to the results
219 obtained using k-means and hierarchical clustering (**Fig.S3**). In summary, our observations
220 demonstrate that ClustALL significantly outperforms classical methodologies regarding population-
221 based robustness.

222

223 Characterization of the five robust stratifications within the PREDICT population

224 After identifying the robust stratifications, we aimed to explore and characterize the distinct clusters
225 observed in each of the five alternative stratifications. These stratifications divided the patients into
226 two clusters, except for stratification 1, which had three clusters. We visually investigated the
227 separation by representing each stratification in a low-dimensional space using the corresponding
228 embeddings derived from the dendrogram depths (**Fig.3A-E**) and the complete dataset (**Fig.S2**).
229 Further exploration revealed that stratification 1 was a subdivision of stratification 2 (**Fig.3F**). We
230 then determined the minimal sets of variables (excluding the cirrhosis severity scores (**Table S1**
231 variables 44 to 48)) with the highest predictive performance in differentiating the clusters for each
232 stratification (**Tables S3-S7**) (27). The different classification approaches were described by 25
233 variables from a total of 74 (**Table S1** variables 1 to 74), with 8 to 12 variables per stratification
234 (**Fig.4A**). Notably, all stratifications included: (i) serum bilirubin concentration (either as a continuous

235 variable or categorized under the term “liver dysfunction”(28)) (ii) International Normalized Ratio
236 (INR) (either as a continuous variable or categorized under the term “coagulation dysfunction” (28);
237 (iii) the number of organ dysfunction or failure. Precipitating events were present in all but one
238 stratification (stratification 3) either as a sum or individually (gastrointestinal bleeding, alcohol-
239 related hepatitis, acute viral hepatitis). Diabetes mellitus was included in two stratifications.
240 Conversely, age, sex, BMI, cause of cirrhosis and lifestyle were present in no or one stratification.
241 Interestingly, stratification 1 and 2 shared almost the same minimal set of variables. Both
242 stratifications identified a group of patients with a severe phenotype attested by low serum sodium,
243 low serum albumin, high serum bilirubin, high INR, high C-Reactive Protein (CRP) and leucocytes, and
244 the number of precipitating events (**Fig.4B**). Hepatic encephalopathy was present in stratification 1
245 but not in 2 (29). A complete statistical characterization of the stratifications is provided in **Tables S3**
246 **to S7**. Considering the clinical implications of the features and the finer classification of the patients,
247 we identified stratification 1 as the most insightful for further exploration in patients with acute
248 decompensation of cirrhosis. Henceforth, in our discussions, we will refer to this specific stratification
249 as '*AD-strat*'.

250

251 *AD-strat* provides prognosis value

252 The *AD-strat* stratification is defined by three subgroups (clusters) of patients with acutely
253 decompensated cirrhosis revealing different clinical characteristics and disease progression. Cluster
254 1 included 306 patients (39.95%) who exhibited the most clinically critical scenario (**Fig.5A, B** and
255 **Table S3**). These individuals had the highest rates of organ dysfunction, clinical events, and
256 precipitating events (**Table S1**). They had a marked acute inflammatory profile (high white blood cell
257 count, and C-reactive protein level) poor liver function (low levels of albumin, and high levels of INR
258 and serum bilirubin) and more hepatocyte injury (higher levels of serum aspartate
259 aminotransferase). Conversely, Cluster 2 (n=118; 15.4%) and Cluster 3 (n=342; 44.6%) had a less
260 severe presentation. The main difference between Cluster 2 and 3 was hepatic encephalopathy,
261 found in 89% of the patients in Cluster 2 and almost no patients in Cluster 3 (**Fig.5A, B** and **Table S2**).
262 Importantly, a significant prognostic value of *AD-strat* was revealed by exploring the cumulative
263 incidence of ACLF and death over 90-day follow-up (**Fig.5C**). Patients in Cluster 1 had poor short-
264 term outcome, with a cumulative incidence of ACLF and death, both by 90 days of 24.1 and 21.5,
265 respectively. While Clusters 2 and 3 had similar risks of ACLF by 90 days (8.6% and 10.2 %,
266 respectively), the risk of death by 90 days was lower for Cluster 2 than Cluster 3 (4.3% vs 10.7%).

267

268 When we compared the clusters identified with ClustALL - exclusively using data obtained at
269 admission - with the groups of patients based on their clinical course (26), we found a statistically
270 significant association (Fisher test, p-value < 0.01) (see **Table S8**). We observed that 61% of patients
271 with pre-ACLF were in Cluster 1 and 48% of patients with stable decompensated were in Cluster 3.

272

273 Reproducibility of the stratification model in an independent cohort

274

275 We assessed the validity of the *AD-strat* model in a large independent prospective multicentre
276 cohort that included 580 patients with acute decompensation of cirrhosis from the Latin-American
277 ACLARA study (30). Using as a reference the PREDICT *AD-strat* clusters, we labelled ACLARA patients
278 using the k-nearest neighbours (kNN) classification algorithm (**Table S9**) (31). The classification
279 model included the 12 predictive variables previously identified in the feature importance analysis

280 (Fig.3B Stratification 1). Importantly, the labelling was consistent and independent of the imputation
281 (Fig.6A), and the distribution of individuals by *AD-strat* clusters within ACLARA closely mirrored that
282 of the PREDICT cohort (Fig.6B). As expected, the clustering of the ACLARA cohort exhibited similar
283 clinical feature patterns to the PREDICT cohort (Fig.6C, Fig.3B Stratification 1). Furthermore, the
284 features describing the subgroups demonstrated statistical significance (Table S10). Finally, we
285 assessed the clinical relevance of the clustering in terms of prognosis, specifically examining the
286 short-term outcomes available in the ACLARA cohort 28 days after hospital admission. Similar to
287 results obtained in the PREDICT cohort, Cluster 1 displayed a bad prognosis for both ACLF and death,
288 while Cluster 3 showed a better prognosis (Fig.6D). In ACLARA, all patients from Cluster 2 were
289 afflicted by hepatic encephalopathy (Table S10) and showed a poor prognosis similar to that of
290 Cluster 1. Ethnicity was homogeneously distributed across clusters (Table S2). In particular, Native
291 Americans represented 21% of Cluster 1, 15% of Cluster 2, and 14% of Cluster 3. Complete
292 information on patient characteristics and short-term outcomes is reported in Supplemental Table
293 9.

294

295 *AD-strat* as a marker for clinical management

296

297 Finally, we investigated the clinical value of the stratification during the follow-up visits of the
298 PREDICT cohort. Based on the PREDICT study design (26), two follow-up visit plans were established
299 depending on the reported disease severity (CLIF-C AD-score) at hospital admission (Fig. 7A). For
300 patients with a CLIF-C AD-score ≥ 50 , the scheduled visits were performed at hospital admission and
301 1, 4, 8 and, 12 weeks after enrolment. For patients with a CLIF-C AD-score < 50 , the scheduled visits
302 were performed only at hospital admission and 1 and 12 weeks after enrolment.

303

304 Of the 766 patients included in the PREDICT study, 688 had at least one follow-up visit. For this
305 subset of patients with available data, we labelled each of them at each follow-up visit using the kNN
306 algorithm (Fig. 7B). This approach allowed an overview of the patient stratification over the entire
307 study duration and revealed the patient flow over time highlighting cluster transitions.

308

309 Consistent with the previous *AD-strat* characterization at hospital admission (Fig.5 and Table S3), we
310 identified more than 50% of patients with CLIF-C AD score ≥ 50 ($n=486$) were classified as Cluster 1,
311 while patients with CLIF-C AD score < 50 ($n=280$) were predominantly classified as Cluster 3 (66.4%)
312 (Fig.7B). Changes in cluster proportions were observed during the patients' follow-up. Stratification
313 changes over time were more pronounced among patients with CLIF-C AD scores ≥ 50 at hospital
314 admission, showing a progressive reduction of patients classified as Cluster 1 (55.8% at HA, 38.8% at
315 week 1, 39.2% at week 4, 25% at week 8, and 17.1% at week 12) and an increase of those classified
316 as Cluster 3 (32.1% at HA, 54.6% at week 1, 50.9% at week 4, 67.9% at week 8, and 74.3% at week
317 12). Additionally, there was a progressive increase in the proportion of patients classified as Cluster
318 3 for those patients with CLIF-C AD-score < 50 at hospital inclusion (66.4% at HA, 83.3% at week 1,
319 and 82.5% at week 12).

320

321 To assess the effectiveness of the *AD-strat* throughout disease progression, we determined its
322 prognostic value in two scenarios: 1) using the stratification at hospital admission, and 2) using the
323 stratification at the last visit reported before the occurrence of any adverse event (we considered
324 any visit between week 1 and 12) or at the end-of-study (EOS) (week 12 visit). A significant

325 difference was observed ($p < 0.001$, Wilcoxon test) when comparing the time window between the
326 visit used in each scenario and the occurrence of adverse events (**Fig. S5**), indicating that in the
327 second scenario, we evaluated patients during a visit much closer to the event.

328

329 Ultimately, the cumulative incidence of ACLF and death as stratified at the last visit demonstrated a
330 more significant separation between clusters compared to patient stratification at hospital admission
331 (**Fig.8**). There was an increase in the incidence for those patients classified as Cluster 1 (18.46% and
332 18.45% at baseline and 28.16% and 26.8% at the last visit for ACLF and death, respectively).
333 Accordingly, the goodness-of-fit parameters indicated an improvement in risk prediction with the last
334 visit stratification, suggesting an enhanced predictive power as the event approached (**Table S11**).

335

336

337 Methods

338 ClustALL Framework

339 Given a set of patients affected by a complex disease with clinical data available, the goal of ClustALL
340 is to identify all the possible alternatives to stratify them that are robust and consistent, even when
341 different parameters or settings are used to generate the stratifications (distance metric, clustering
342 algorithm, and the number of imputations).

343

344 Input data

345 ClustALL accepts both binary and numerical clinical variables as input. Categorical features are
346 transformed using a one-hot encoder method. A minimum of two features is required, but including
347 more features would lead to more precise clustering. It is important to note that increasing the
348 number of features may also increase the computation time.

349

350 Step 1. Data Complexity Reduction

351 In this step, highly correlated features are replaced by a reduced set of variables that account for
352 their variability. To that end:

353 Step 1.1. Dendrogram. Hierarchical clustering is applied to the data, resulting in a
354 dendrogram where variables are grouped based on similarity (32). The depth of each branch
355 represents the distance between the groups of variables. All the possible depths of the dendrogram
356 are extracted, and the sets of variables beneath each depth are stored as *Depth*.

357 Step 1.2. Preprocessing. Principal Component Analysis (PCA) is computed for each set of
358 variables corresponding to each *Depth*, and the first three principal components are stored in a new
359 matrix (*Embedding*) (33). For sets that contain only one variable, the variable itself is stored to
360 generate the replacement matrix. This results in a complexity-reduced data set (*Embedding*) for each
361 considered *Depth*. A subset of depths can be considered when the number of variables is too large.

362

363 Step 2 Stratification Process

364 In this step, ClustALL calculates and pre-evaluates stratifications for each *Embedding*. For each
365 *Embedding*, the dissimilarity between patients' pairs is computed using correlation-based distance
366 and Gower dissimilarity metric, resulting in two distance matrices. Clustering algorithms are then
367 applied (34–36) depending on the distance used: k-means and hierarchical clustering for correlation
368 distance matrices, and k-medoids and hierarchical clustering for the Gower distance matrix.
369 Throughout all experiments, five different cluster numbers are evaluated $k \in \{2, 3, 4, 5, 6\}$. The
370 optimal number of clusters for each strategy is determined based on the consensus from three
371 different measures of clustering internal validation: the sum-of-squares based index or WB-ratio, the
372 Dunn index, and the average silhouette width (37,38). The objective is to group patients with
373 comparable data while ensuring that patients in separate clusters are as dissimilar as possible from
374 those in other clusters. As the output for this step, a stratification is derived for each combination
375 denoted as "*embedding + distance metric + clustering method*".

376

377 Step 3 Consensus-based stratifications

378 Step 3.1. Population-based robustness. A data-driven threshold is used to define
379 population-based robust subgroups or clusters. For each resulting stratification from the previous
380 step, cluster-wise stability is computed by bootstrapping the dataset 1,000 times and calculating the

381 Jaccard similarity index to the originally defined clusters (39). Stratifications with less than 85%
382 stability (**Fig.S4**) are excluded based on data distribution. The remaining stratifications are denoted
383 as *Strat_{fit}*.

384 Step 3.2. Jaccard distance is applied to compute distances between the population-based
385 robust stratifications. Then, to identify parameter-based robust clusters (where a minor modification
386 in parameter selection provides a similar result), ClustALL considers those combinations that are part
387 of a group of stratifications (green squares in Consensus-based stratifications step in **Fig.1**). Then, as
388 initial criteria, that can be modified by the user, centroids from each "combination group" are
389 selected as parameter-based robust stratifications (coloured green squares in Consensus-based
390 stratifications step in **Fig.1**). The outcome can be none, one, or multiple ways to stratify the
391 population robustly. In the current analysis, we considered parameter-based robust representatives:
392 centroids of a combination group that includes at least 5 population-based robust stratifications.
393

394 **ClustALL enables input data with missing values**

395 ClustALL can be adapted to work with missing data (**Fig.S1**). To that end, the ClustALL method is
396 modified as follows:

397 Step 1 adaptation. First, a dendrogram and its associated depths are computed considering the
398 original dataset with missing values. The original dataset is then imputed 1,000 times with the MICE
399 algorithm to ensure the results are not derived from a single imputation (40). For each *Depth*
400 *previously calculated* and each imputed dataset, the Data Complexity Reduction step is applied.

401 Step 2 adaptation. Step 2.1 is computed for each combination of depth, distance metric, clustering
402 algorithm and each *Embedding* derived from an imputed dataset. The selection of the optimal
403 number of clusters is based on the consensus from cluster internal validation and the mode of the
404 imputed datasets for each corresponding embedding. Afterward, a distance matrix (D_{mat}) between
405 individuals is obtained by computing how often two individuals are assigned to the same cluster in
406 each imputation (**Fig.S1**). Then, D_{mat} calculates a final stratification score using correlation-based
407 distance and h-clust. In our experience, limited optimization is required here because summarizing
408 the stratification over all imputations separately strengthens what is observed in each imputed
409 dataset. Extra care will be required only in cases where imputations may differ significantly. After
410 this modification, the method follows as previously described (**Fig.S1**).
411

412 **Data source**

413 The data utilized in this study were obtained from two independent multicentre studies: the
414 European PREDICT cohort and the Latin-American ACLARA cohort, conducted as part of the European
415 Project DECISION (26,41). Both cohorts collected various measures including clinical,
416 pharmacological, biomarker, and outcome data from patients with acute decompensation of
417 cirrhosis upon hospital admission and during follow-up visits. The follow-up period was 90 days for
418 the PREDICT cohort and 28 days for the ACLARA cohort. To be eligible for the present study, patients
419 were required to have acute decompensation of cirrhosis upon hospital admission, with available
420 information on short-term outcomes, drug intake, and available biological samples. Ultimately, 766
421 patients from the PREDICT cohort and 580 patients from the ACLARA cohort and 74 features
422 (continuous and categorical) were included in the analysis. The features included demographic
423 information, clinical and laboratory data, medical history, risk factors, and cirrhosis scores at hospital
424 admission, with missing values accounting for less than 30% (**Table S1**). To avoid bias from missing
425 data, imputation was performed with 1,000 iterations using the Multivariate Imputation by Chained

426 Equations (MICE) method (21).

Demographics	Age, sex, height [†] , weight [†] , BMI, ethnicity (Black or African American, Asian, White, other)
Cause of cirrhosis	Alcohol, viral, alcohol + viral, NASH, cryptogenic, other
Main reason for hospitalization	Ascites, hepatic encephalopathy, gastrointestinal bleeding, spontaneous bacterial peritonitis, other infection
Manifestations at admission	Clinical Events (ascites, hepatic encephalopathy, gastrointestinal bleeding, acute kidney injury, bacterial infection, acute alcoholic-steatohepatitis, acute viral hepatitis, hepatocellular carcinoma), number of clinical events (the sum of clinical events), number of precipitating events (the sum of precipitating events: proven bacterial infection, acute alcoholic-steatohepatitis, CLIF-C AD > 50), organ dysfunction (liver, renal, cerebral, coagulation, cardiac, respiratory), number of organ dysfunctions (the sum of organ dysfunctions), organ failure (liver, cerebral, coagulation, cardiac, respiratory), number of organ failures (the sum of organ failures)
Cirrhosis Severity Scores	Child-Pugh, CLIF-C AD, CLIF-C OF, MELD, MELDNA
Medical history	History of diabetes, history of hypertension, history of previous decompensations
Lifestyle	Alcohol, active alcohol consumption [†] , tobacco
Laboratory variables (measured in serum)	Alanine transaminase, aspartate aminotransferase, albumin, bilirubin (total), gamma-GT, C-reactive protein, sodium, potassium, glucose [†] , hemoglobin, hematocrit, creatinine, white blood cell count, lymphocytes, monocytes, neutrophils, INR (International Normalized Ratio), platelet, SpO2 (%), SpO2/FiO2 Ratio

427 **Table 1. Complete list of input features.** Patient characteristics included in the analysis: demographics,
428 cause of cirrhosis, main reason for hospitalization, manifestations at admission, cirrhosis severity
429 scores, medical history, lifestyle and laboratory variables. [†]Variables not included in ACLARA cohort.
430

431 **CLustALL comparison to different clustering methodologies**

432 A comparison was conducted between the ClustALL framework and classical clustering algorithms.
433 Stratification was performed on 1,000 imputed datasets using classical k-means and hierarchical
434 methodologies with k values of 2 and 3, considering that ClustALL robust stratifications comprised
435 two or three patient subgroups. Bootstrapping was performed for the classical clusters to evaluate
436 cluster-wise stability (39). The resulting stability was compared to ClustALL stability through the
437 Kolmogorov–Smirnov test. Moreover, the clinical utility of the various stratifications was assessed by
438 examining the clinical insights obtained from the different clusters.
439

440 Statistical Methods

441 All analyses were performed in the R Computing Environment version 4.0.3 (42).

442

443 Descriptive statistics

444 Descriptive characteristics of the PREDICT and ACLARA study populations were reported as means
445 with standard deviations for continuous variables and proportions of patients for categorical
446 variables.

447 Feature Analysis

448 The identification of the minimal-size predictive signatures with maximal predictive power leading to
449 each stratification was performed using the *fbcd.reg* function with default hyperparameters from
450 the 'MXM' R package (27).

451 Parametric Tests

452 Differences between clusters in the PREDICT and in the ACLARA cohorts were assessed using one-
453 way ANOVA for continuous variables, while binary variables were tested with the chi-square test.
454 The association between the PREDICT clusters identified with ClustALL - exclusively using data
455 obtained at admission - with the groups of patients based on their clinical course (26), was tested
456 with the Fisher test.

457 Stratification model reproducibility

458 *AD-strat* model was validated in and in a separate cohort of patients with acute decompensation of
459 cirrhosis from the ACLARA cohort and in PREDICT follow-up time points. For this purpose, the kNN
460 model was trained on the PREDICT *AD-strat* cluster labels based on the signatures previously defined
461 as most predictive in the feature analysis. The K parameter was selected based on accuracy, the area
462 under the curve (AUC), error rate (ER), false positives (FP), and false negatives (FN) (**Table S7**). After
463 applying the KNN algorithm, the target data (ACLARA cohort and PREDICT follow-up) was labelled
464 based on the majority votes from the K nearest neighbours and imputed datasets.

465 Survival Analysis

466 Cumulative incidences of ACLF development and liver-related death were estimated using the
467 cumulative incidence function of the 'survival' R Package. Liver transplantation was considered a
468 competing event. A p-value lower than 0.05 with Benjamini and Hochberg (BH) adjustment was
469 considered statistically significant.

470

471 Longitudinal analysis and model evaluation

472 All PREDICT patients with ≥ 1 post-baseline assessment ($n = 688$) were included in longitudinal
473 outcomes analyses for a period of 90 days after hospital admission. Sankey diagrams were
474 generated to show the patients' transfers among the *AD-strat* clustering, liver transplant, ACLF
475 development, death and survival status. The predictive power of the stratification models at
476 different time points in the PREDICT cohort was evaluated using BIC, AIC, Concordance, and
477 Likelihood ratio goodness-of-fit parameters (41).

478

479 Discussion

480

481 Traditional patient classification methods based on outcome prediction scores may overlook
482 important heterogeneity factors. To address this issue, we developed ClustALL, a computational
483 pipeline that handles various clinical data challenges and identifies robust patient stratifications. We
484 applied ClustALL to a cohort of patients with acute decompensation of cirrhosis, identifying five
485 robust stratifications with prognostic value.

486

487 Optimal patient stratification is required to sustain the precision medicine revolution occurring in
488 the clinical setting (43). While significant progress has been made in classification problems,
489 particularly in domains like single-cell transcriptomic analysis (44,45), unsupervised clustering of
490 patients based on clinical information is still in the developmental stage (7,46). Notably, the existing
491 challenges in clinical stratification are often addressed using ad-hoc solutions that consider mixed
492 data types, missing values, or highly correlated variables. However, no comprehensive method
493 currently exists that addresses all these challenges. To overcome the aforementioned idiosyncrasies
494 in clinical data, we have developed a novel computational framework named ClustALL. Significantly,
495 beyond addressing existing challenges, ClustALL improves over other existing methodologies by
496 allowing the identification of more than one robust stratification within a given population. Clinical
497 data is complex and allows for multiple uses and "multiple interpretations" that *may* result in several
498 valid groupings (47). Indeed, the concept of "multiple interpretations" arises from how variables are
499 utilized in the clustering process and has been a subject of research in the early 21st century (48).
500 Another distinctive feature of ClustALL is the consistency of the resulting representative
501 stratifications even when limited modifications in the clustering parameter settings are applied. In
502 the context of biological data, such as gene expression data, this property has already been already
503 defined as the "propensity of a clustering algorithm to maintain output coherence over a range of
504 settings"(49). Interestingly, this definition has been applied in the study of exposome and
505 pregnancy-related mortality in the United States (50). In summary, we believe that ClustALL
506 represents one of the initial necessary steps towards incorporating two necessary features into
507 clinical stratification: parameter-based robustness and the identification of more than one
508 stratification.

509

510 To assess the effectiveness of ClustALL, we applied it as a proof-of-concept in a cohort of patients
511 with acutely decompensated cirrhosis considering data collected at hospital admission. Such an
512 attempt to apply stratification to patients with cirrhosis has never been conducted. The stratification
513 we set up differs from the scores developed and routinely used in patients with cirrhosis (*e.g.*, MELD,
514 MELD-Na, Child-Pugh, CLIF-C-AD) both in terms of design and use. Indeed, all these scores were built
515 using a follow-up endpoint (usually death) in patients receiving therapies. These scores are helpful to
516 identify patients at high risk of poor outcomes, but they do not fully capture the heterogeneity of
517 the patients at admission for several reasons: (a) some features explaining patients heterogeneity
518 might not have an independent prognostic value, either because the prognostic information they
519 carry is contained in other variables, or because therapies administered to patients during their
520 follow-up blunt their impact; (b) a similar survival rate does not imply similar pathophysiological
521 mechanisms. For instance, in PREDICT, clusters 2 and 3 have a similar rate of ACLF, while they
522 strongly differ with regard to the prevalence of hepatic encephalopathy. The stratification presented
523 here is not intended to guide clinical bedside decisions or to replace a prognostic score but rather to

524 identify homogeneous patient populations at hospital admission. This stratification could base the
525 development of future clinical trials including more homogeneous patient populations. In this
526 regard, to make this stratification easily accessible to all, we developed an online calculator and
527 application available at <https://decision-for-liver.eu/for-scientists/clustall-web-application/>. The
528 purely data-driven approach and the development and independent validation of the stratifications
529 in large prospective multicentre European and Latin American cohorts strengthen our results.

530

531 Through this analysis, as a first step, we identified five alternative stratifications for patients with
532 acute decompensation of cirrhosis. Interestingly, all stratifications included markers of impaired liver
533 function, namely serum bilirubin and INR, but also the number of organ dysfunction or failure, and
534 all but one included precipitating events. This emphasizes that these features are crucial when
535 designing a clinical trial including patients with acute decompensation of cirrhosis. On the contrary,
536 some features like age, sex, BMI, cause of cirrhosis and lifestyle were present in no or only one
537 stratification suggesting that these features are not key when designing a clinical trial. The
538 stratification we selected (*AD-strat*) provided a more granular resolution by allowing the
539 identification of three subgroups of patients. In this stratification, diabetes mellitus is taken into
540 account. While it is known that diabetes is an independent risk factor for cirrhosis decompensation
541 (51,52), the role of diabetes once acute decompensation has happened has been overlooked so far.
542 This place of diabetes is quite unique since causes of cirrhosis, comorbidities or lifestyle were not
543 part of the key features of *AD-strat*. Hepatic encephalopathy strongly impacted the categorization of
544 patients with acutely decompensated cirrhosis. Notably, 89% and 100% of the patients in Cluster 2
545 from the PREDICT and ACLARA cohorts, respectively, presented hepatic encephalopathy at the time
546 of hospital admission. This may explain the intermediate prognosis observed in patients within
547 Cluster 2, as hepatic encephalopathy is recognised by its fluctuating nature and potential
548 reversibility (53,54). The dynamic nature of hepatic encephalopathy may also explain why Cluster 2
549 was not a static group over time (55).

550

551 Furthermore, tracking patients over time using *AD-strat* labelling allowed for dynamic and improved
552 identification of patients at high risk of adverse events in the PREDICT cohort. These results highlight
553 the ability of the ClustALL not only to stratify patients using baseline characteristics but also that the
554 use of *AD-strat* labelling over time is able to improve this prediction.

555

556 Although our study showed promising results, it is important to acknowledge some limitations.
557 Firstly, our stratification was based solely on routinely available clinical data at hospital admission,
558 which may not provide a comprehensive view of patients' conditions. Future studies should extend
559 our findings with biological data, ideally derived from multiomic analyses. Moreover, it is crucial to
560 consider that the predictive power in the ACLARA cohort was only assessed at 28 days due to the
561 study design.

562

563 In conclusion, this study introduces a novel unsupervised clustering framework, ClustALL, able to
564 overcome the limitations of previously available stratification methods. ClustALL is available as
565 OpenSource (https://github.com/TranslationalBioinformaticsUnit/ClustALL_AD/). When applied to
566 the setting of acute decompensation of cirrhosis, ClustALL enhanced our understanding of patients'
567 heterogeneity emphasizing the importance of liver function and the number of organ dysfunctions
568 or failures, precipitating events, and conversely the limited role of age, sex, BMI, cause of cirrhosis

569 and lifestyle at this stage of the liver disease. The selected stratification, *AD-strat*, might be a useful
570 tool to better design future clinical trials by including more homogeneous patient populations.

571 Data availability

572 Researchers who provide a methodology sound proposal can apply for the data, as far as the
573 proposal is in line with the research consented by the patients. These proposals should be requested
574 through <https://www.clifresearch.com/decision/Home.aspx>. Data requestors will need to sign a data
575 transfer agreement. The code to generate the ClustALL method is available on GitHub, at
576 https://github.com/TranslationalBioinformaticsUnit/ClustALL_AD/.

577

578 Acknowledgements

579

580 This project has received funding from the European Union’s Horizon 2020 research and innovation
581 program under grant agreement No 847949.

582

583 The study was supported by the European Foundation for the Study of Chronic Liver Failure (EF-Clif).
584 The EF-Clif is a nonprofit private organization. The EF-Clif receives unrestricted donations from Cellex
585 Foundation and Grifols. EF-Clif is partner, contributor and coordinator in several EU Horizon 2020
586 program projects. JT was appointed as visiting Professor in EF-Clif for the execution of the study by a
587 grant from Cellex Foundation. The funders had no influence on study design, data collection and
588 analysis, decision to publish or preparation of the manuscript. The fact that EF-CLIF.

589

590 Jonel Trebicka was supported by the German Research Foundation (DFG) project ID 403224013 –
591 SFB 1382 (A09), by the German Federal Ministry of Education and Research (BMBF) for the DEEP-
592 HCC project and by the Hessian Ministry of Higher Education, Research and the Arts (HMWK) for the
593 ENABLE and ACLF-I cluster projects. The MICROB-PREDICT (project ID 825694), DECISION (project ID
594 847949), GALAXY (project ID 668031), LIVERHOPE (project ID 731875), and IHMCSA (project ID
595 964590) projects have received funding from the European Union’s Horizon 2020 research and
596 innovation program. The manuscript reflects only the authors’ views, and the European Commission
597 is not responsible for any use that may be made of the information it contains. The funders had no
598 influence on study design, data collection and analysis, decision to publish, or preparation of the
599 manuscript.

600

601 N.P.P was funded by a Ramón y Cajal fellow (RYC2021-032197-I) from the
602 MCIN/AEI/10.13039/501100011033 and European Union “NextGenerationEU”/PRTR and by a Juan
603 de la Cierva-formation fellow (FJC2019-042304-I) from the Spanish Ministry of Science and
604 Innovation (MCIN).

605

606 P-E.R.’s research laboratory is supported by the Foundation pour la Recherche Médicale (FRM
607 EQU202303016287), “Institut National de la Santé et de la Recherche Médicale” (ATIP AVENIR), the
608 “Agence Nationale pour la Recherche” (ANR-18-CE14-0006-01, RHU QUID-NASH, ANR-18-IDEX-0001,
609 ANR-22-CE14-0002) by « Émergence, Ville de Paris », by Fondation ARC and by the European Union’s
610 Horizon 2020 research and innovation programme under grant agreement No 847949.

611

612 Bibliography

613

- 614 1. Almendro V, Kim HJ, Cheng YK, Gonen M, Itzkovitz S, Argani P, et al. Genetic and
615 phenotypic diversity in breast tumor metastases. *Cancer Res.* 2014 Mar
616 1;74(5):1338–48.
- 617 2. Kotelnikova E, Kiani NA, Abad E, Martinez-Lapiscina EH, Andorra M, Zubizarreta
618 I, et al. Dynamics and heterogeneity of brain damage in multiple sclerosis. *PLoS*
619 *Comput Biol.* 2017 Oct 1;13(10).
- 620 3. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. Disease progression
621 and treatment response in data-driven subgroups of type 2 diabetes compared with
622 models based on simple clinical features: an analysis using clinical trial data. *Lancet*
623 *Diabetes Endocrinol.* 2019 Jun 1;7(6):442–51.
- 624 4. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C, Abraldes JG, et al.
625 Baveno VII – Renewing consensus in portal hypertension. Vol. 76, *Journal of*
626 *Hepatology.* Elsevier B.V.; 2022. p. 959–74.
- 627 5. Moreau R, Jalan R, Gines P, Pavesi M, Angeli P, Cordoba J, et al. Acute-on-Chronic
628 Liver Failure Is a Distinct Syndrome That Develops in Patients With Acute
629 Decompensation of Cirrhosis. *Gastroenterology.* 2013 Jun;144(7):1426-1437.e9.
- 630 6. Cerezo Cerezo J& ALC. Population stratification: a fundamental instrument used for
631 population health management in Spain: good practice brief. . World Health
632 Organization Regional Office for Europe
633 <https://apps.who.int/iris/handle/10665/345586>. 2018;
- 634 7. Moral TT, Sanchez-Niubo A, Monistrol-Mula A, Gerardi C, Banzi R, Garcia P, et al.
635 Methods for Stratification and Validation Cohorts: A Scoping Review. Vol. 12,
636 *Journal of Personalized Medicine.* MDPI; 2022.
- 637 8. Horne E, Tibble H, Sheikh A, Tsanas A. Challenges of clustering multimodal clinical
638 data: Review of applications in asthma subtyping. Vol. 8, *JMIR Medical Informatics.*
639 *JMIR Publications Inc.;* 2020.
- 640 9. Wang H, Donoho D, Kuppler C, Loftus TJ, Jr UG, Copyright frai, et al. Phenotype
641 clustering in health care: A narrative review for clinicians.
- 642 10. Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of
643 clustering techniques and developments. *Neurocomputing.* 2017 Dec 6;267:664–81.
- 644 11. Cismondi F, Fialho AS, Vieira SM, Reti SR, Sousa JMC, Finkelstein SN. Missing
645 data in medical databases: Impute, delete or classify? *Artif Intell Med.* 2013
646 May;58(1):63–72.
- 647 12. Rodríguez AH, Ruiz-Botella M, Martín-Loeches I, Jimenez Herrera M, Solé-Violan
648 J, Gómez J, et al. Deploying unsupervised clustering analysis to derive clinical
649 phenotypes and risk factors associated with mortality risk in 2022 critically ill
650 patients with COVID-19 in Spain. *Crit Care.* 2021 Dec 1;25(1).
- 651 13. Curtis JR, Weinblatt M, Saag K, Bykerk VP, Furst DE, Fiore S, et al. Data-Driven
652 Patient Clustering and Differential Clinical Outcomes in the Brigham and Women’s
653 Rheumatoid Arthritis Sequential Study Registry. *Arthritis Care Res (Hoboken).* 2021
654 Apr 1;73(4):471–80.
- 655 14. Pudjihartono N, Fadason T, Kempa-Liehr AW, O’Sullivan JM. A Review of Feature
656 Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in*
657 *Bioinformatics.* 2022 Jun 27;2.
- 658 15. Hennig C. What are the true clusters? *Pattern Recognit Lett.* 2015 Oct 15;64:53–62.
- 659 16. Lopez-Martinez-Carrasco A, Juarez JM, Campos M, Canovas-Segura B. A
660 methodology based on Trace-based clustering for patient phenotyping. *Knowl Based*
661 *Syst.* 2021 Nov 28;232.
- 662 17. Chalancon G, Kruse K, Babu MM. Clustering Coefficient. In: *Encyclopedia of*
663 *Systems Biology.* New York, NY: Springer New York; 2013. p. 422–4.

- 664 18. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of
665 cluster analysis. *J Comput Appl Math*. 1987 Nov;20:53–65.
- 666 19. Adam SP, Alexandropoulos SAN, Pardalos PM, Vrahatis MN. No Free Lunch
667 Theorem: A Review. In 2019. p. 57–82.
- 668 20. Milligan GW, Cooper MC. An examination of procedures for determining the
669 number of clusters in a data set. *Psychometrika*. 1985 Jun;50(2):159–79.
- 670 21. Steinley D, Brusco MJ. Choosing the number of clusters in K-means clustering.
671 *Psychol Methods*. 2011;16(3):285–97.
- 672 22. Altman N, Krzywinski M. Clustering. *Nat Methods*. 2017 Jun 30;14(6):545–6.
- 673 23. Kitano H. Towards a theory of biological robustness. *Mol Syst Biol*. 2007 Jan
674 18;3(1).
- 675 24. Yu H, Chapman B, Di Florio A, Eischen E, Gotz D, Jacob M, et al. Bootstrapping
676 estimates of stability for clusters, observations and model selection. *Comput Stat*.
677 2019 Mar 28;34(1):349–72.
- 678 25. Lu Y, Phillips CA, Langston MA. A robustness metric for biological data clustering
679 algorithms. *BMC Bioinformatics*. 2019 Dec 24;20(S15):503.
- 680 26. Trebicka J, Fernandez J, Papp M, Caraceni P, Laleman W, Gambino C, et al. The
681 PREDICT study uncovers three clinical courses of acutely decompensated cirrhosis
682 that have distinct pathophysiology. *J Hepatol*. 2020 Oct 1;73(4):842–54.
- 683 27. Tsagris M, Tsamardinos I. Feature selection with the R package MXM. *F1000Res*.
684 2018 Jan 1;7:1505.
- 685 28. Arroyo V, Moreau R, Jalan R. Acute-on-Chronic Liver Failure. *New England Journal*
686 *of Medicine*. 2020 May 28;382(22):2137–45.
- 687 29. Stewart CA, Malinchoc M, Kim WR, Kamath PS. Hepatic encephalopathy as a
688 predictor of survival in patients with end-stage liver disease. *Liver Transplantation*.
689 2007 Oct;13(10):1366–71.
- 690 30. Farias AQ, Vilalta AC, Zitelli PM, Pereira G, Goncalves LL, Torre A, et al. Genetic
691 Ancestry, Race, and Severity of Acutely Decompensated Cirrhosis in Latin America.
692 *Gastroenterology* [Internet]. 2023 May; Available from:
693 <https://linkinghub.elsevier.com/retrieve/pii/S0016508523008065>
- 694 31. Deng Z, Zhu X, Cheng D, Zong M, Zhang S. Efficient kNN classification algorithm
695 for big data. *Neurocomputing*. 2016 Jun 26;195:143–8.
- 696 32. Schonlau M. Visualizing non-hierarchical and hierarchical cluster analyses with
697 clustergrams. Vol. 19, *Computational Statistics*. 2004.
- 698 33. Ringnér M. What is principal component analysis? [Internet]. Vol. 26, *NATURE*
699 *BIOTECHNOLOGY*. 2008. Available from:
700 <http://www.nature.com/naturebiotechnology>
- 701 34. Hummel M, Edelmann D, Kopp-Schneider A. Clustering of samples and variables
702 with mixed-type data. *PLoS One*. 2017 Nov 1;12(11).
- 703 35. Zhang Z, Murtagh F, Poucke S Van, Lin S, Lan P. Hierarchical cluster analysis in
704 clinical research with heterogeneous study population: Highlighting its visualization
705 with R. *Ann Transl Med*. 2017 Feb 1;5(4).
- 706 36. Arora P, Deepali, Varshney S. Analysis of K-Means and K-Medoids Algorithm for
707 Big Data. In: *Physics Procedia*. Elsevier B.V.; 2016. p. 507–12.
- 708 37. Liu Y, Li Z, Xiong H, Gao X, Wu J. Understanding of internal clustering validation
709 measures. In: *Proceedings - IEEE International Conference on Data Mining, ICDM*.
710 2010. p. 911–6.
- 711 38. Zhao Q, Fränti P. WB-index: A sum-of-squares based index for cluster validity. *Data*
712 *Knowl Eng*. 2014;92:77–89.

- 713 39. Hennig C. Cluster-wise assessment of cluster stability. *Comput Stat Data Anal*. 2007
714 Sep 15;52(1):258–71.
- 715 40. Van Buuren S, Groothuis-Oudshoorn K. *Journal of Statistical Software mice: Multivariate Imputation by Chained Equations in R* [Internet]. Vol. 45. 2011.
716 Available from: <http://www.jstatsoft.org/>
- 717 41. Farias AQ, Curto Vilalta A, Momoyo Zitelli P, Pereira G, Goncalves LL, Torre A, et al. Genetic Ancestry, Race, and Severity of Acutely Decompensated Cirrhosis in Latin America. *Gastroenterology*. 2023 Sep;165(3):696–716.
- 718 42. R Core Team. *R: A language and environment for statistical .* Vienna, Austria: R Foundation for Statistical Computing; 2021.
- 719 43. Johnson KB, Wei WQ, Weeraratne D, Frisse ME, Misulis K, Rhee K, et al. Precision Medicine, AI, and the Future of Personalized Health Care. Vol. 14, *Clinical and Translational Science*. Blackwell Publishing Ltd; 2021. p. 86–93.
- 720 44. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. Vol. 20, *Nature Reviews Genetics*. Nature Publishing Group; 2019. p. 273–82.
- 721 45. Qi R, Ma A, Ma Q, Zou Q. Clustering and classification methods for single-cell RNA-sequencing data. Vol. 21, *Briefings in Bioinformatics*. Oxford University Press; 2019. p. 1196–208.
- 722 46. Coombes CE, Liu X, Abrams ZB, Coombes KR, Brock G. Simulation-derived best practices for clustering clinical data. *J Biomed Inform*. 2021 Jun 1;118.
- 723 47. Müller E, Günemann S, Färber I, Seidl T. Discovering multiple clustering solutions: grouping objects in different views of the data. In: *Proceedings - International Conference on Data Engineering*. 2012. p. 1207–10.
- 724 48. Hu J, Pei J. Subspace multi-clustering: a review. Vol. 56, *Knowledge and Information Systems*. Springer London; 2018. p. 257–84.
- 725 49. Lu Y, Phillips CA, Langston MA. A robustness metric for biological data clustering algorithms. *BMC Bioinformatics*. 2019 Dec 24;20.
- 726 50. Harville EW, Grady SK, Langston M, Juarez PJ, Vilda D, Wallace ME. The public health exposome and pregnancy-related mortality in the United States: a high-dimensional computational analysis. *BMC Public Health*. 2022 Dec 1;22(1).
- 727 51. Elkrief L, Rautou PE, Sarin S, Valla D, Paradis V, Moreau R. Diabetes mellitus in patients with cirrhosis: clinical implications and management. *Liver International*. 2016 Jul;36(7):936–48.
- 728 52. Paternostro R, Jachs M, Hartl L, Simbrunner B, Scheiner B, Bauer D, et al. Diabetes impairs the haemodynamic response to non-selective betablockers in compensated cirrhosis and predisposes to hepatic decompensation. *Aliment Pharmacol Ther*. 2023 Jul 31;
- 729 53. Romero-Gómez M, Montagnese S, Jalan R. Hepatic encephalopathy in patients with acute decompensation of cirrhosis and acute-on-chronic liver failure. *J Hepatol*. 2015 Feb;62(2):437–47.
- 730 54. Ferenci P. Hepatic encephalopathy. *Gastroenterol Rep (Oxf)*. 2017 May;5(2):138–47.
- 731 55. Higuera-de-la-Tijera F, Velarde-Ruiz Velasco JA, Raña-Garibay RH, Castro-Narro GE, Abdo-Francis JM, Moreno-Alcántar R, et al. Current vision on diagnosis and comprehensive care in hepatic encephalopathy. *Revista de Gastroenterología de México (English Edition)*. 2023 Apr;88(2):155–74.
- 732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760

761 Figure Legends

762

763 **Figure 1. Schematic overview of the different steps of ClustALL approach (best viewed in colour).**

764 ClustALL takes clinical variables as input. First, data complexity is reduced by grouping the features
765 into a dendrogram, assessing the resulting depths, and using Principal Component Analysis (PCA)
766 (green panel). The output is an embedding for each possible depth. Then, stratification is
767 computed considering the combination of different distance measures, clustering techniques, and
768 cluster numbers (K) (purple panel). In the final step, non-robust stratifications are filtered, and the
769 centroids derived from computing Jaccard (coloured green squares) similarity among the robust
770 stratifications (green squares) are considered the final representatives of the stratifications (red
771 panel).

772 **Figure 2. Summary of the outputs from the different steps of the ClustALL framework when**

773 **applied to the PREDICT cohort (N=766).** Input data comprised 74 clinical features with less than
774 30% missing values. The analysis utilized 1,000 imputed datasets. The Data Complexity Reduction
775 step (green) was applied to 72 depths of the 1,000 imputed datasets. The Stratification Process
776 step (purple) considered various clustering combinations resulting in 288 stratifications. After
777 bootstrapping, 144 robust stratifications remained. Finally, in the Consensus-based Stratification
778 step (red), five groups of robust stratifications (red squares) were identified, and the centroid was
779 selected from each group as the final stratifications (red coloured squares).

780 **Figure 3. Principal Component projection of the ClustALL robust stratifications based on the**

781 **embedding associated with each stratification.** (A-E). Low-dimension representation of the robust
782 stratifications after applying the ClustALL framework to the PREDICT cohort. For each one of the 5
783 robust stratifications identified by ClustALL, the Principal Component Analysis of the Embeddings
784 corresponding to the specific dendrogram depth associated with the stratification is shown. The x
785 (Dim1) and y (Dim2) axes represent the first and second principal components respectively, which
786 are linear combinations of the original variables. (F). The overlap between the clusters in
787 stratifications 1 and 2 shows that stratification 1 is a subdivision of stratification 2.

788 **Figure 4. Overview of the variables driving the ClustALL stratifications.** (A). Heatmap with the

789 minimal set of variables required to describe the 5 different stratifications, accounting for 25 out of
790 74 input variables. (B). Heatmaps of the minimal set of patient characteristics per stratification. The
791 heatmap colour scale depends on the data type. In the case of binary variables, the value indicates
792 the percentage of patients with such binary characteristics, e.g., the presence of Diabetes Mellitus.
793 For continuous variables, the colour scale represents a scaled value from the highest cluster mean
794 (100.0) to the lowest cluster mean (0.0), e.g., Albumin and CRP. Abbreviations: ASH = Acute
795 Alcoholic-Steatohepatitis, AST = Aspartate aminotransferase, CL= Cluster, CRP = C-Reactive Protein,
796 HE = Hepatic encephalopathy, HCC = Hepatocellular Carcinoma, INR = International normalized
797 ratio, WBC = White blood cell counts.

798

799 **Figure 5. Clinical overview of the AD-strat derived clusters in the PREDICT cohort. (A, B).**

800 Distribution of the highest predictive performance-related patient characteristics among AD-strat
801 clusters; (A) categorical variables, (B) numerical variables. C) Cumulative incidence of ACLF (left)
802 and death (right) according to the AD-strat clustering in PREDICT cohort considering 90 days after
803 hospital admission, with the number of patients at risk per cluster (Transplantation counted as a
804 competing risk to death). Abbreviations: AST = Aspartate aminotransferase, CRP = C- Reactive
805 Protein, INR = International normalized ratio, WBC = White blood cell counts.

806 **Figure 6. Reproducibility of the AD-strat model in the ACLARA cohort.** (A) Distribution of the

807 labels in the ACLARA cohort after applying the kNN model 1,000 times. (B) Proportion of patients

808 distributed in the 3 clusters in the PREDICT and the ACLARA cohorts. (C) Heatmap of patient
809 characteristics per cluster in the ACLARA cohort. Bars on the right show the colour scale
810 representing the proportion with each binary characteristic, such as diabetes. Continuous
811 variables, such as bilirubin, represent a scaled value from the highest cluster mean (1.0) to the
812 lowest cluster mean (0.0). (D) Cumulative incidence of ACLF (up) and death (down) according to
813 the AD-strat clustering in ACLARA cohort considering 28 days after hospital admission, with the
814 number of patients at risk per cluster (Transplantation counted as a competing risk to death).
815 Abbreviations: AST = Aspartate aminotransferase, CRP = C- Reactive Protein, INR = International
816 normalized ratio, WBC = White blood cell counts.

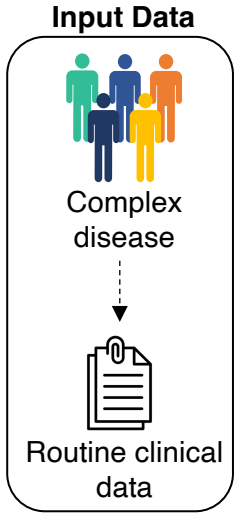
817

818 **Figure 7. Distribution and transition of the AD-strat derived clusters at different visits in the**
819 **PREDICT cohort.** (A). Schematic representation of PREDICT study design. Two follow-up visit plans
820 were defined according to the reported disease severity (CLIF-C AD-score) at hospital admission
821 (red). The information about the occurrence of any adverse event (liver transplant, ACLF or death)
822 during the whole visit plan or the absence of events at the end of the study was tracked (blue). (B)
823 Sankey plots show the cluster label of each patient over the follow-up visits. The follow-up flows of
824 patients with CLIF-C AD ≥ 50 at hospital admission (up) and CLIF-C AD < 50 at hospital admission
825 (down) are shown. The distribution of the patients assessed at each follow-up visit per cluster is
826 shown as frequency and proportion on the top of the Sankey representations. The accumulated
827 frequency and proportion of adverse events at each follow-up visit respecting the whole cohort
828 (for CLIF-C AD ≥ 50 , $n=486$; for CLIF-C AD < 50 , $n=280$) are shown on the bottom of the Sankey
829 representations. Reported event/eos, shows the status of a patient at the “end of the study”:
830 patients with a reported event or patients with no reported event.

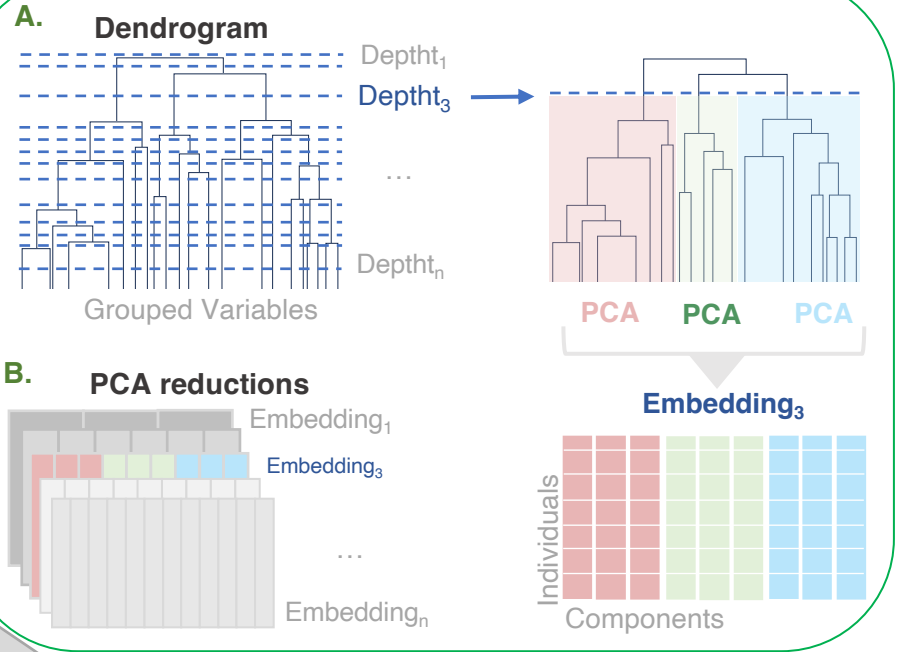
831

Figure 1

ClustALL



1. Data Complexity Reduction



2. Stratification Process

For each embedding:

Distance metric + Clustering Method

- Correlation
- Gower
- H-Clust
- K-means
- K-medoids

Evaluation of the optimal number of clusters

$K \in 2, 3, 4, 5, 6$ (WB-ratio, Dunn and Silhouette)

Stratifications

(Embedding + Distance + Clustering)

3. Consensus-based Stratifications

1. Population-based robust stratifications by bootstrapping
2. Parameter-based robust stratifications

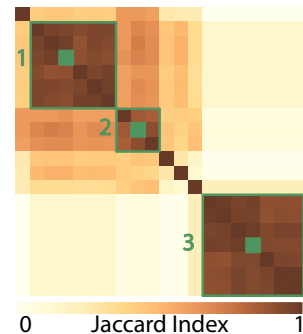
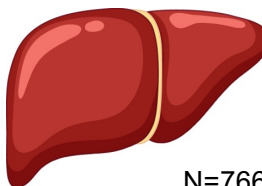


Figure 2

PREDICT cohort



N=766

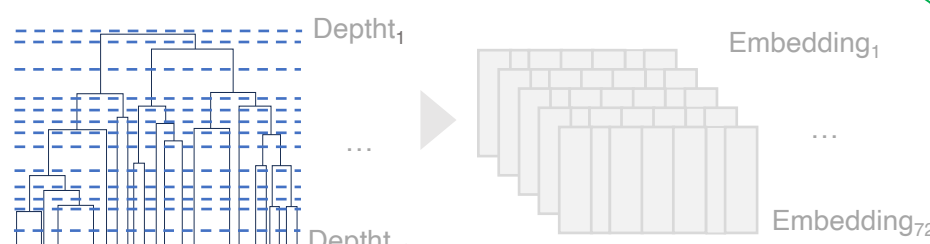
Input Data (74 variables)

- Demographics
- Clinical examination and laboratory data
- Cirrhosis severity scores
- Medical history
- Lifestyle

(Features with >30% NAs excluded)

1. Data Complexity Reduction

1000x
Imputation
From original dataset

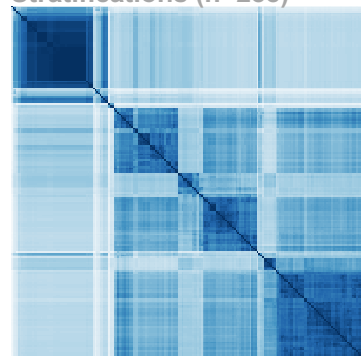


72,000 Embeddings

2. Stratification Process

Imputation + Embedding + Distance & Clustering

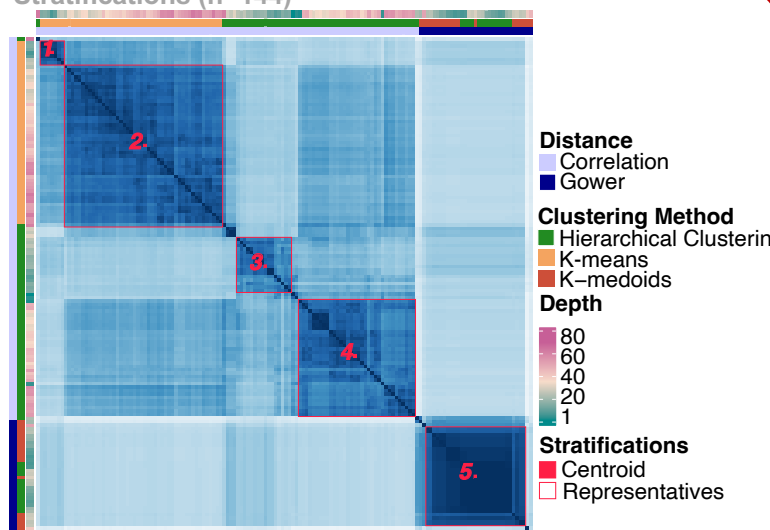
Stratifications (n=288)



0 Jaccard Index 1

3. Consensus-based Stratifications

Stratifications (n=144)



0 Jaccard Index 1

Distance

- Correlation
- Gower

Clustering Method

- Hierarchical Clustering
- K-means
- K-medoids

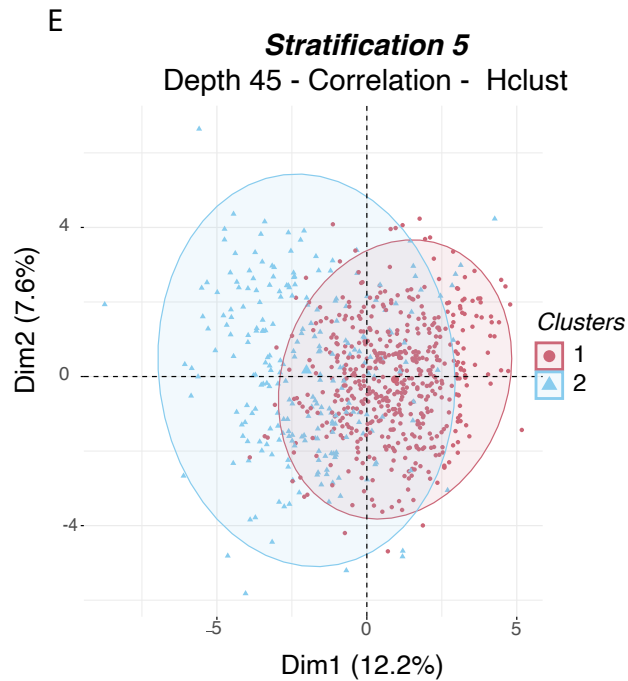
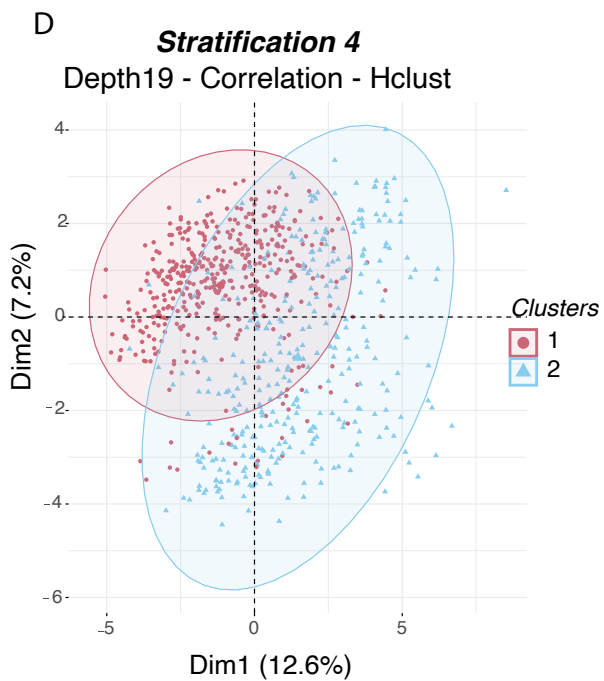
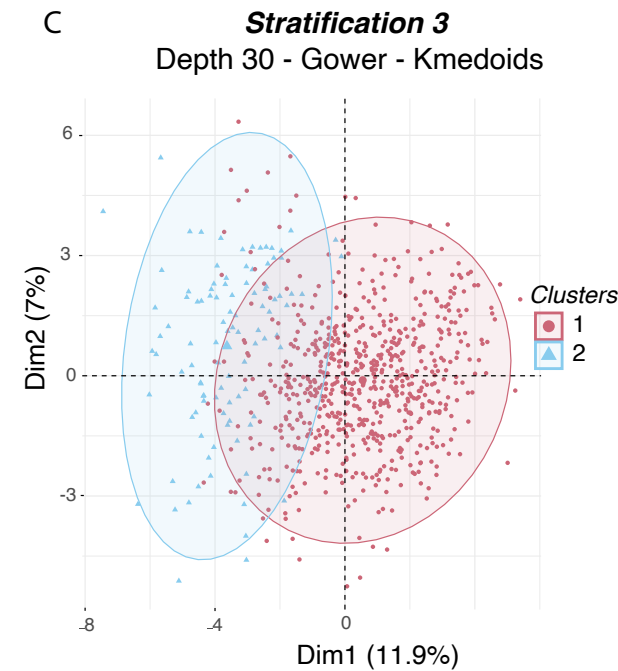
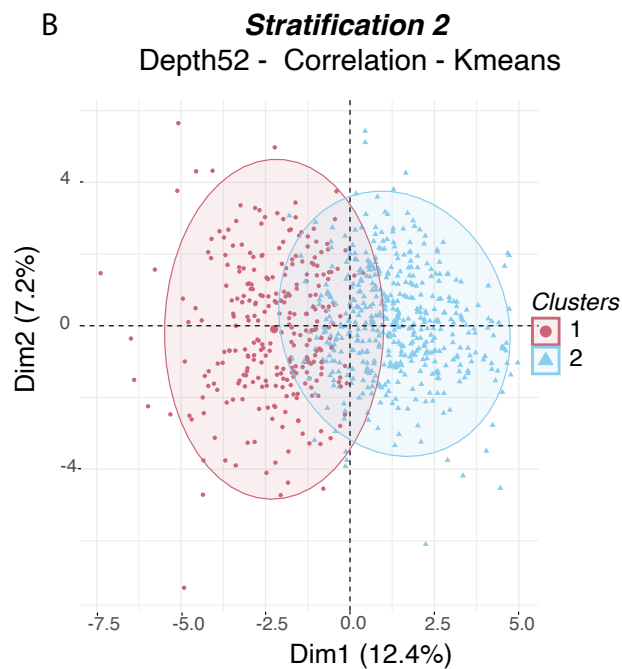
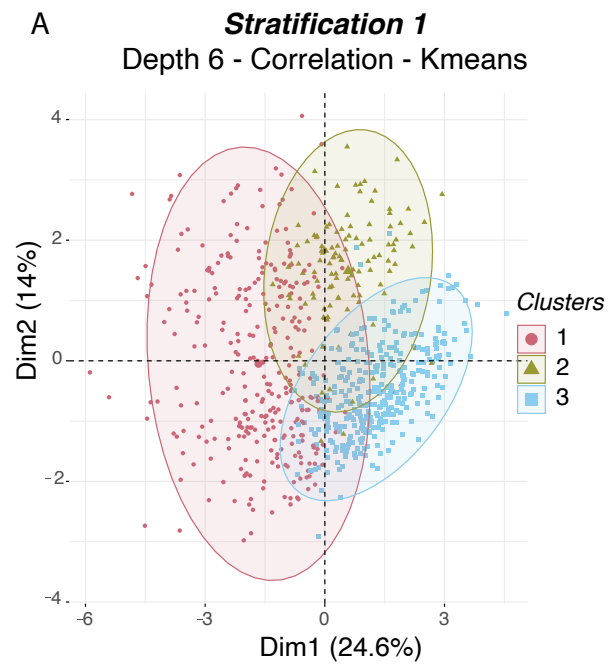
Depth

- 80
- 60
- 40
- 20
- 1

Stratifications

- Centroid
- Representatives

Figure 3

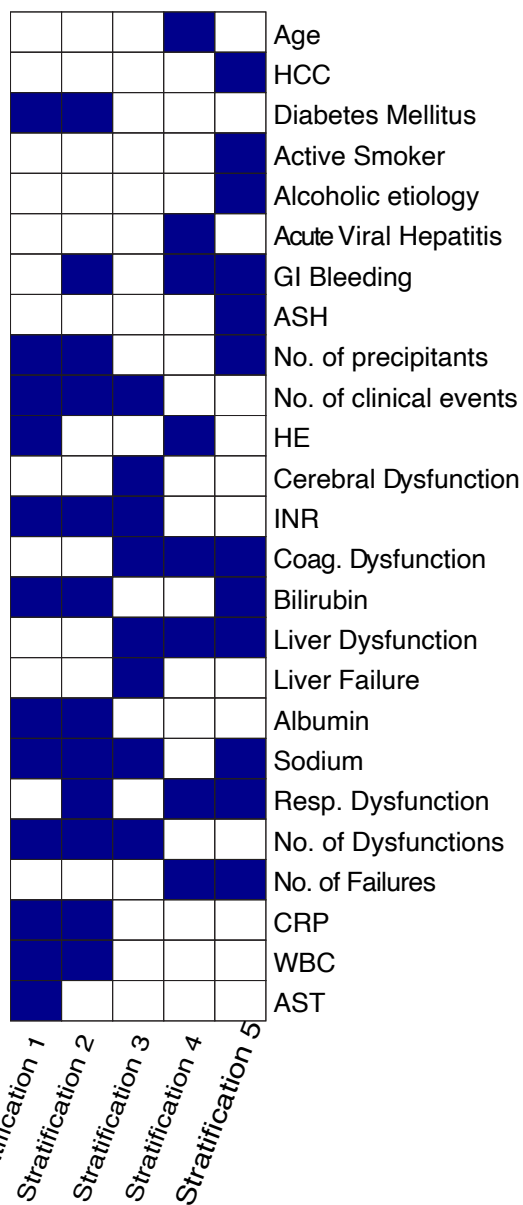


F

	Stratification 2	
Stratification 1	Cluster 1	Cluster 2
Cluster 1	227	29
Cluster 2	4	114
Cluster 3	2	340

Figure 4

A



B

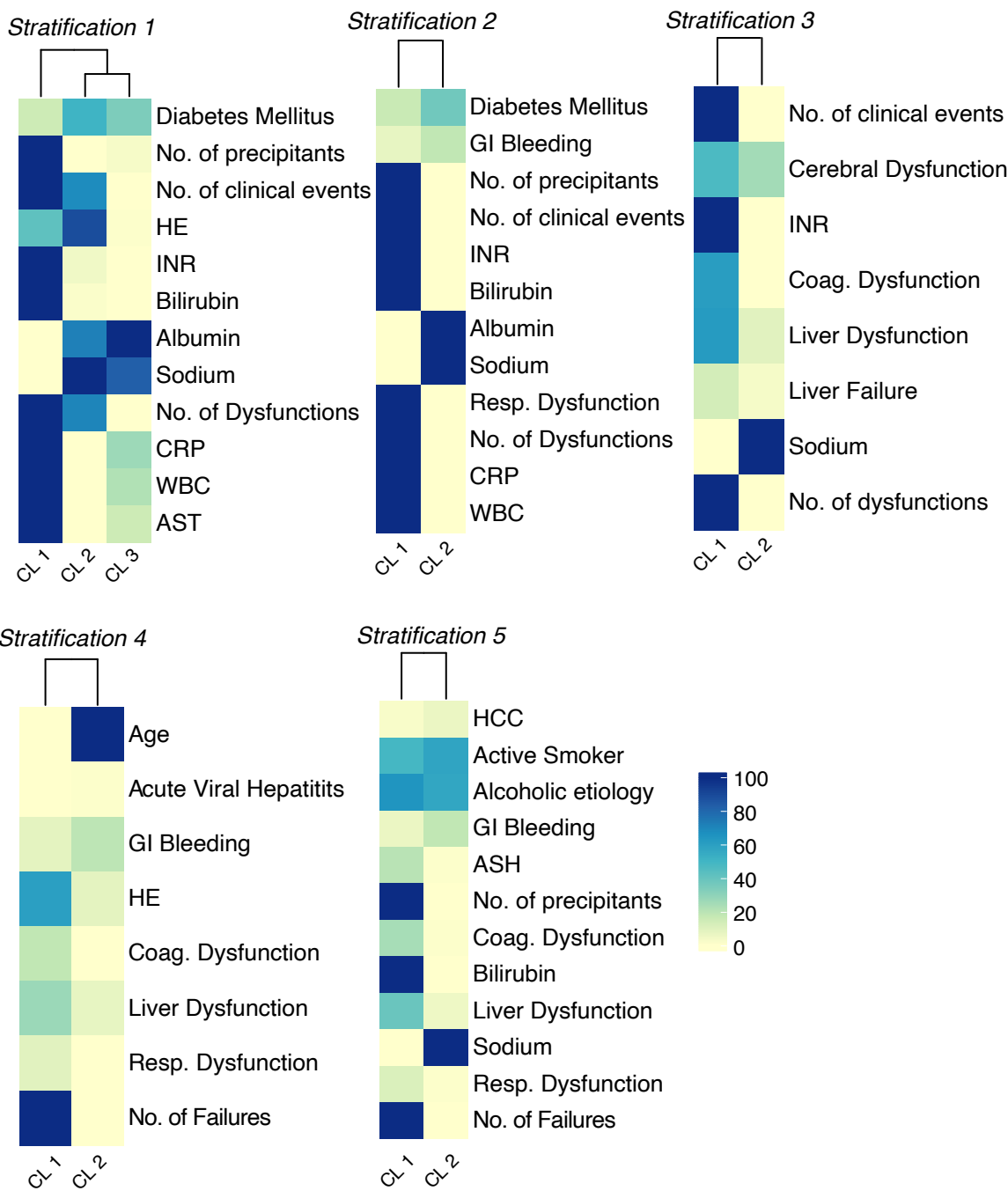


Figure 5

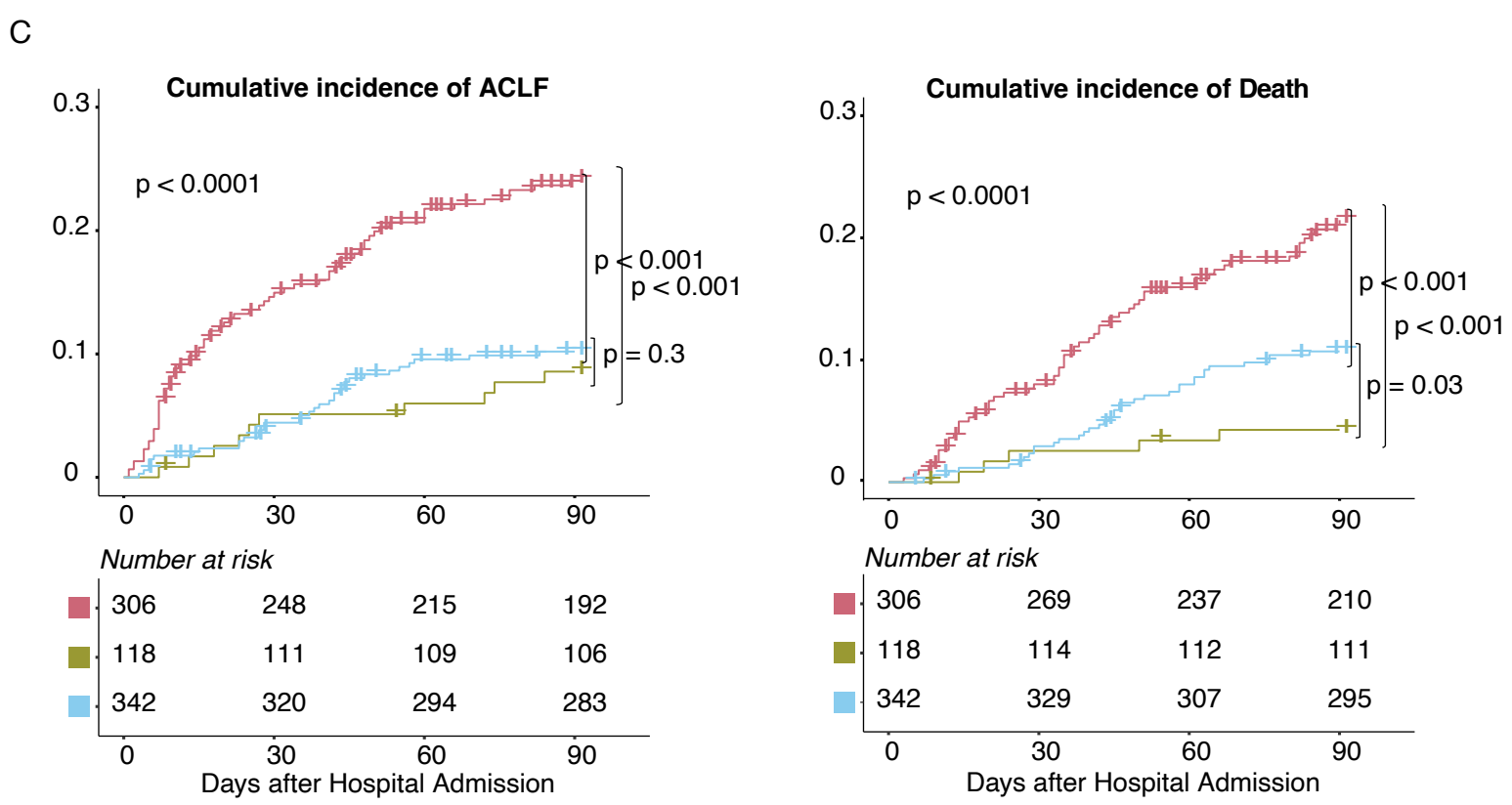
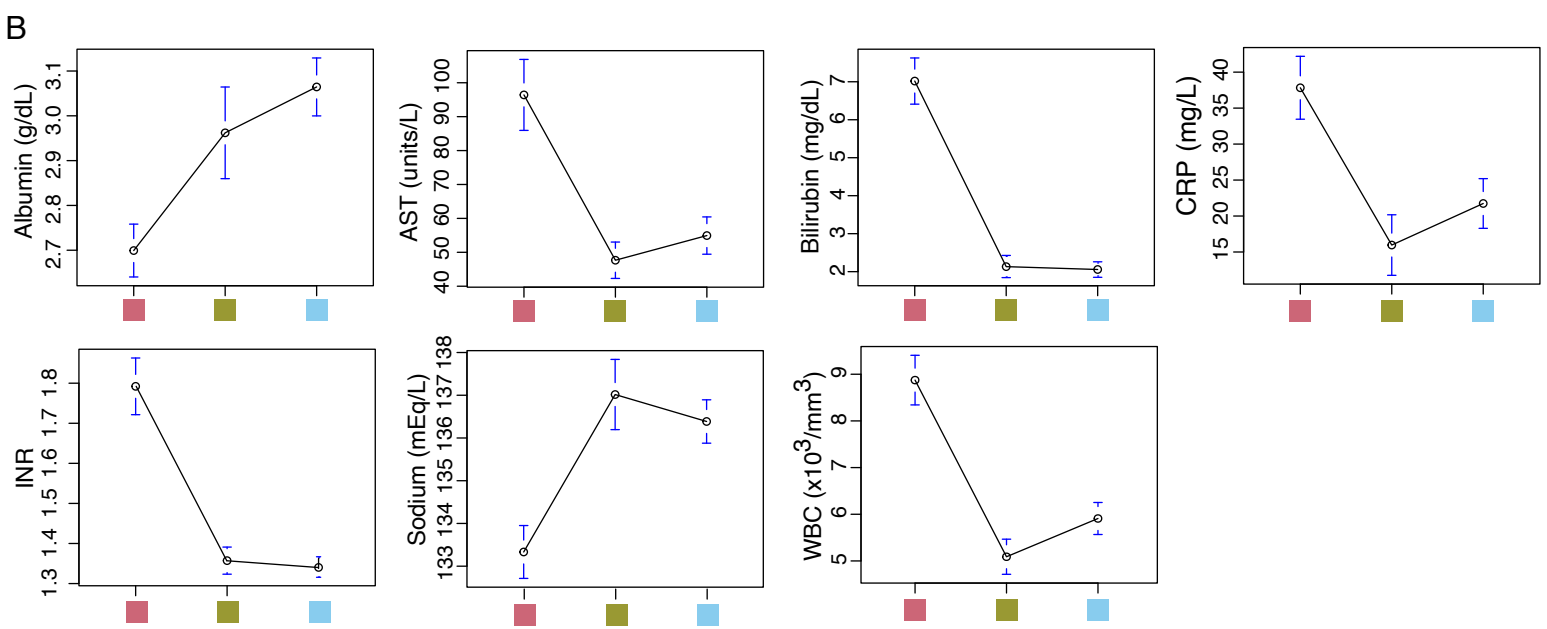
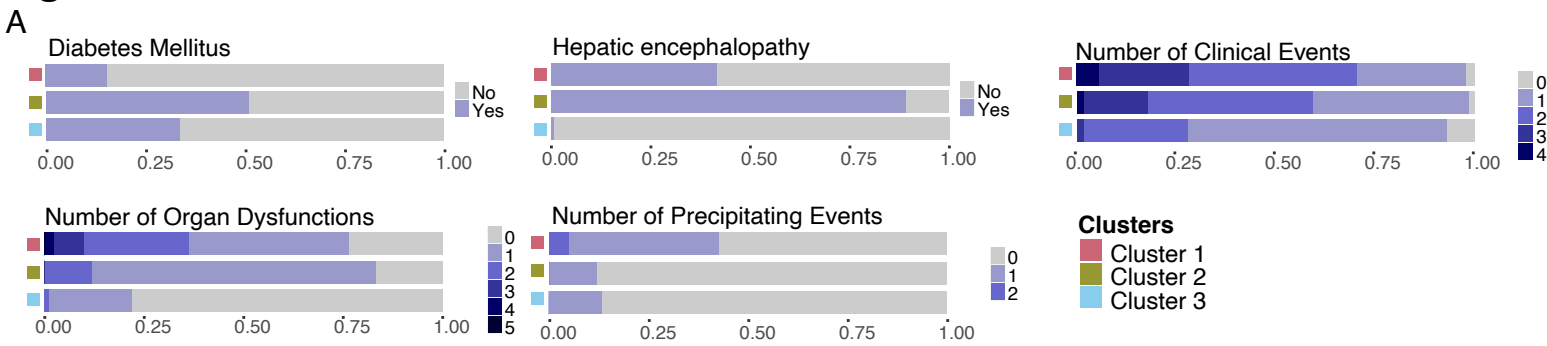
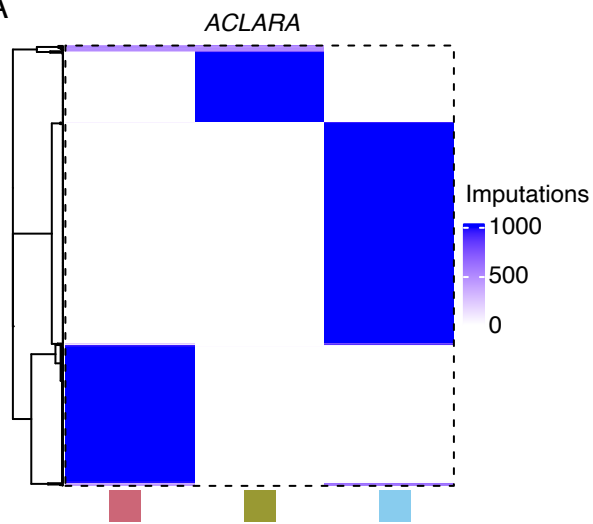
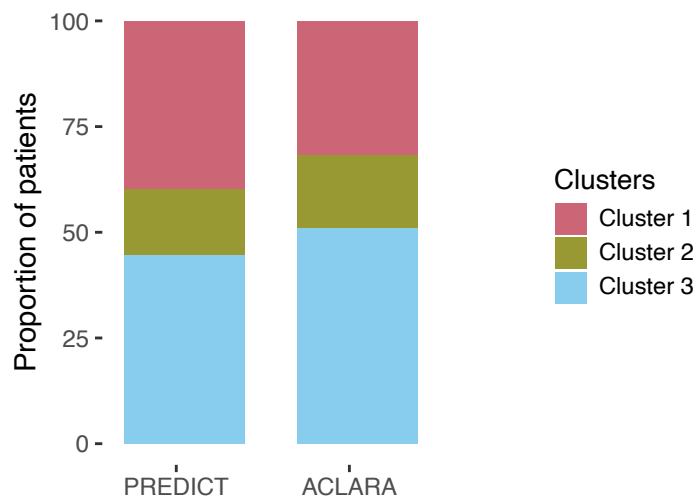


Figure 6

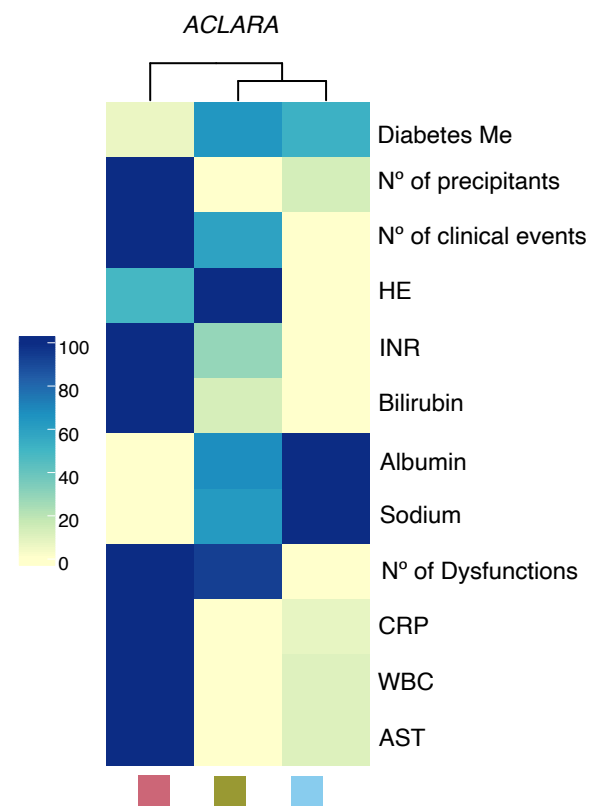
A



B



C



D

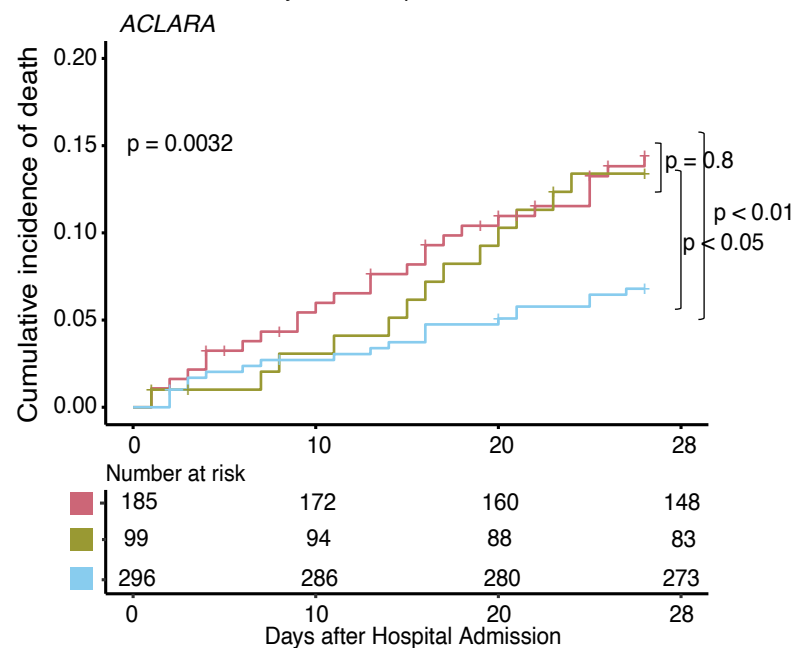
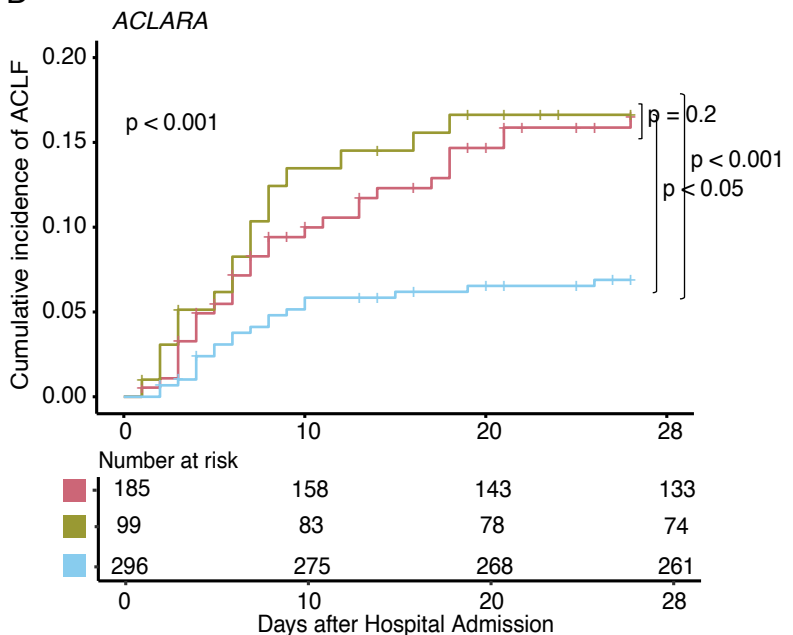
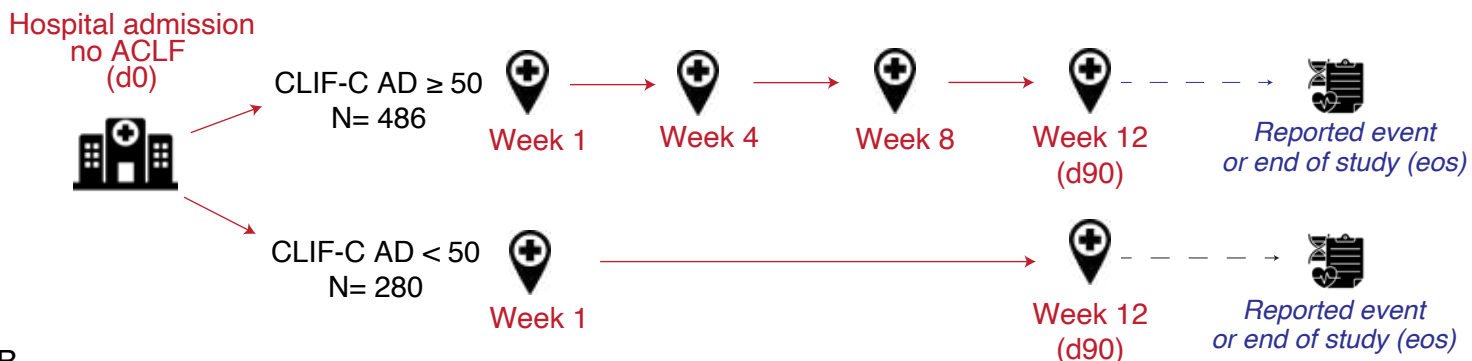


Figure 7

A



B

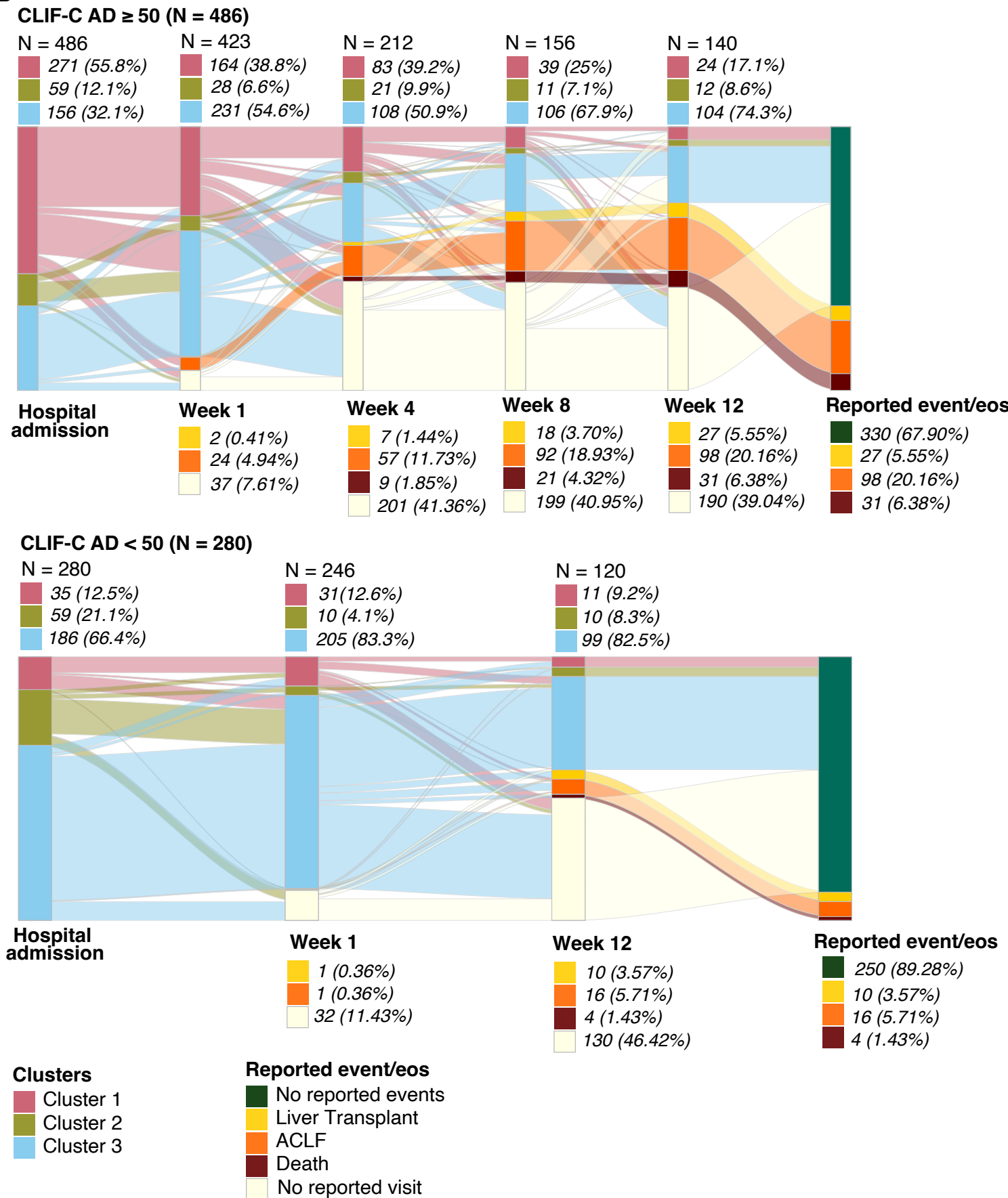
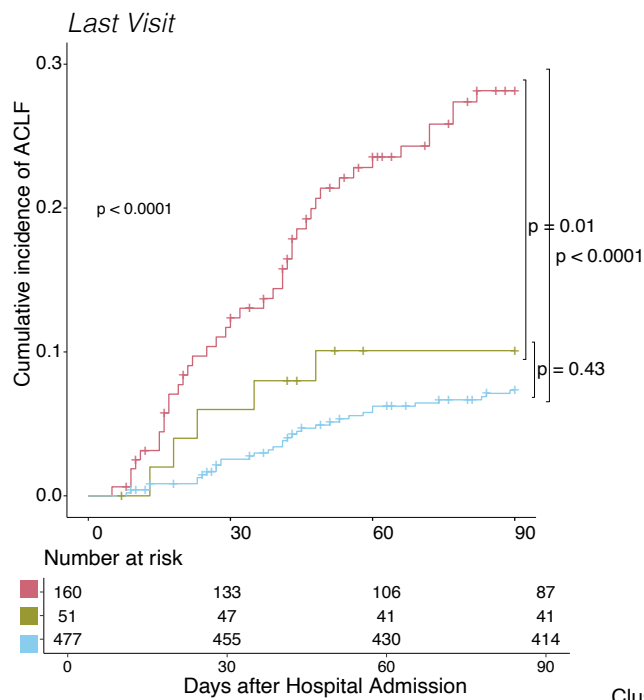
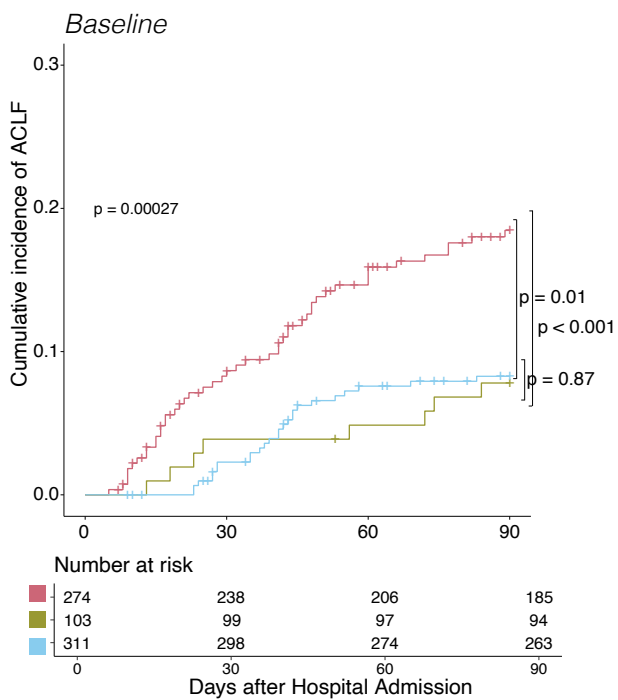


Figure 8

A

Risk of ACLF



Clusters

- Cluster 1
- Cluster 2
- Cluster 3

B

Risk of death

