Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An exploratory routine-laboratory renal dysfunction-coagulopathy phenotype identifies high-risk ICU patients with pneumonia: MIMIC- IV derivation and fixed-centroid Chinese external validation

YuLiang Zhang
Longhua Hospital, Shanghai University of Traditional Chinese Medicine
ShaoHeng Zhang
Longhua Hospital, Shanghai University of Traditional Chinese Medicine
JiaYue Li
Longhua Hospital, Shanghai University of Traditional Chinese Medicine
XinLei Wang
Longhua Hospital, Shanghai University of Traditional Chinese Medicine
Qi Gao
Longhua Hospital, Shanghai University of Traditional Chinese Medicine
Ran Chen
Longhua Hospital, Shanghai University of Traditional Chinese Medicine
YiYi Zhang
eezh@msn.com

Longhua Hospital, Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: pneumonia, intensive care unit, MIMIC-IV, phenotype, renal dysfunction, coagulopathy, external validation, routine laboratory tests

Posted Date: August 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10336848/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Critically ill adults with pneumonia are clinically heterogeneous. Conventional severity scores summarize acute organ dysfunction but may not define clinically interpretable patient states. We investigated whether routine laboratory tests obtained early after ICU admission could identify reproducible phenotypes among ICU patients with pneumonia and whether these phenotypes could be externally validated in a Chinese ICU cohort.

Methods

We conducted a retrospective cohort study using MIMIC-IV as the derivation cohort and a single-center Chinese ICU cohort for fixed-centroid external assignment. Adult patients with pneumonia-related diagnoses, first ICU stay, and ICU length of stay of at least 24 hours were included in MIMIC-IV. The external cohort included adults with severe pneumonia admitted to the ICU of Longhua Hospital between January 1, 2024 and March 31, 2025. Eleven routine laboratory measures obtained during the first ICU day represented hematologic status, coagulation, renal function and electrolyte balance. K-means solutions from k = 2 to k = 6 were evaluated. The exploratory k = 3 solution used 20 random initializations, and repeated multistart analyses assessed sensitivity to initialization. Local patients were assigned without re-clustering by applying MIMIC-derived imputation, standardization and nearest-centroid rules. The primary outcome was in-hospital mortality. Secondary outcomes included CRRT/RRT and ICU length of stay.

Results

The MIMIC-IV derivation cohort included 16,933 ICU stays, with 3,687 in-hospital deaths. The external cohort included 412 patients, with 120 deaths. A three-phenotype exploratory solution was selected. Phenotype 1 showed a relatively preserved hematologic-renal profile and the lowest risk. Phenotype 2 showed cytopenic and electrolyte-shift features with intermediate risk. Phenotype 3 was characterized by renal dysfunction, coagulation abnormality and electrolyte derangement. In MIMIC-IV, mortality increased from 16.9% in phenotype 1 to 22.6% in phenotype 2 and 32.2% in phenotype 3. CRRT use increased from 2.2% to 5.2% and 22.5%, respectively. In the Chinese cohort, phenotype 3 also showed the highest mortality (50.0%) and CRRT use (63.9%). Compared with phenotype 1, phenotype 3 remained associated with mortality after adjustment for age, sex and SOFA score (OR 1.193, 95% CI 1.068–1.332). Bootstrap analysis identified phenotype 3 as the highest-mortality and highest-CRRT cluster in 99% of iterations. In 30 repeated analyses using 20 random initializations per analysis, all k = 3 solutions had an adjusted Rand index above 0.99 relative to the reference solution.

Conclusions

Early routine laboratory tests identified an exploratory renal dysfunction-coagulopathy high-risk phenotype among ICU patients with pneumonia. This phenotype was associated with higher mortality

and markedly higher CRRT use in MIMIC-IV and showed directionally consistent fixed-centroid assignment in a Chinese ICU cohort. These results support further multicentre and prospective evaluation of routine-laboratory phenotyping; they do not establish phenotype-guided treatment.

Highlights

• Early routine laboratory tests identified reproducible phenotypes among ICU patients with pneumonia.

• The most robust finding was a renal dysfunction–coagulopathy high-risk phenotype.

• This phenotype had higher mortality and markedly higher CRRT use in MIMIC-IV.

• Fixed-centroid assignment showed directionally consistent external validation in a Chinese ICU cohort.

• Routine-laboratory phenotyping may be more transportable than biomarker-heavy frameworks.

Background

Pneumonia is a major reason for ICU admission and remains associated with substantial mortality, organ support and prolonged critical care resource use. In the ICU, pneumonia is not a uniform syndrome. Patients admitted under the same diagnostic label may differ substantially in hematologic reserve, coagulation status, renal function, electrolyte balance and overall physiologic stability. This heterogeneity complicates clinical stratification and may limit the interpretability of models that focus only on predicting mortality.

Traditional severity scores such as SOFA and APACHE-based systems summarize the degree of acute physiologic derangement [11, 12], but they do not necessarily define clinically interpretable subgroups within pneumonia. Similarly, many machine-learning studies in ICU pneumonia have focused on predicting death, respiratory support or length of stay. Although prediction is useful, it may not explain whether distinct laboratory-defined patient states exist or whether these states can be recognized across different healthcare systems.

Data-driven phenotyping offers an alternative approach. Rather than estimating a single risk score, phenotyping identifies groups of patients with similar clinical profiles. This strategy aligns with the broader movement to redefine critical illness through biologically and clinically meaningful patient states rather than syndromic labels alone [2, 3]. In critical care, phenotypes and subphenotypes have been used to describe heterogeneity in ARDS and sepsis, where inflammatory or molecular subphenotypes are associated with different outcomes and potential treatment responses [4–7].

Pneumonia-specific subphenotyping is also emerging. Recent studies have identified clinically meaningful subphenotypes in hospital-acquired pneumonia and community-acquired pneumonia using multicohort or prospective designs [8, 9]. However, many phenotype frameworks rely on biomarkers, prospective research measurements or disease-specific cohorts that may be difficult to reproduce in

routine electronic health record data. Biomarkers such as C-reactive protein, procalcitonin, D-dimer, fibrinogen or natriuretic peptides may have biological value, but they are not measured consistently across hospitals and databases.

MIMIC-IV provides a large, detailed electronic health record resource for critical care research [1]. However, transportability across healthcare systems remains a major concern for data-driven models [10]. We therefore focused on routine laboratory tests that are widely available in ICU practice and can be harmonized across MIMIC-IV and a local Chinese ICU dataset. We hypothesized that early routine laboratory data could identify reproducible phenotypes among critically ill adults with pneumonia and that a high-risk phenotype characterized by renal dysfunction and coagulation abnormality would show consistent adverse outcome associations across datasets.

Methods

Study design and data sources

This retrospective cohort study used MIMIC-IV version 3.1 as the derivation cohort and a Chinese single-center ICU dataset as the external validation cohort. MIMIC-IV contains de-identified critical care data from Beth Israel Deaconess Medical Center [1]. The external validation cohort was derived from the hospital information system of the Department of Critical Care Medicine, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China.

The analysis proceeded in four steps. First, an adult ICU pneumonia cohort was identified in MIMIC-IV. Second, early routine laboratory variables were extracted and used to derive phenotypes. Third, phenotype-outcome associations were assessed in MIMIC-IV. Fourth, local patients were assigned to the MIMIC-derived phenotypes and used for external validation (Fig. 1). The study was structured according to principles for observational reporting and transparent model reporting [13,14].

Study population

In MIMIC-IV, eligible patients were adults aged 18 years or older with a pneumonia-related ICD diagnosis during the same hospitalization, a first ICU stay and an ICU length of stay of at least 24 hours. Pneumonia-related diagnoses were identified by linking diagnoses_icd to d_icd_diagnoses and selecting descriptions containing "pneumonia" or "pneumonitis due to inhalation". This rule identified 157 ICD-9/10 codes and 43,092 hospital admissions. The first_icu_stay field in the MIMIC-IV derived icustay_detail table defined the first ICU stay. The complete code list and executable cohort query are provided in the supplementary reproducibility files.

The external validation cohort included adult patients admitted to the ICU of Longhua Hospital between January 1, 2024 and March 31, 2025 who met local diagnostic criteria for severe pneumonia. Patients were required to have ICU length of stay of at least 24 hours and available early laboratory and outcome

data. Patient identifiers and exact admission dates were removed from the analytic dataset before analysis.

Phenotype-defining variables

Candidate variables were selected according to clinical relevance, cross-database availability and early-window missingness. The final core phenotype set included white blood cell count, hemoglobin, platelet count, prothrombin time, activated partial thromboplastin time, blood urea nitrogen, creatinine, sodium, potassium, chloride and calcium. These variables represent hematologic status, coagulation, renal function and internal-environment balance.

Variables directly representing outcomes or treatments, including mortality, ICU length of stay, mechanical ventilation and renal replacement therapy, were not used for clustering. SOFA and APACHE/APS scores were used for phenotype characterization and adjusted outcome analyses, not for phenotype discovery. These scores were treated as established severity measures rather than clustering inputs [11, 12].

Data preprocessing

For MIMIC-IV, laboratory variables were obtained from the first ICU day using the derived first_day_lab table. To represent clinically adverse early values consistently, the analysis used the maximum WBC, PT, APTT, BUN, creatinine and potassium values and the minimum hemoglobin, platelet, sodium, chloride and calcium values. For the local cohort, the corresponding early ICU measurements were harmonized to MIMIC-compatible units. Hemoglobin was converted from g/L to g/dL, urea to BUN in mg/dL, creatinine from umol/L to mg/dL and calcium from mmol/L to mg/dL.

Missing values in phenotype-defining variables were imputed using MIMIC-derived medians. Continuous variables were standardized using MIMIC-derived means and standard deviations. The same MIMIC imputation and standardization parameters were applied to the local cohort before external phenotype assignment.

Phenotype discovery and external assignment

K-means clustering was performed in the MIMIC-IV derivation cohort using the 11 standardized core laboratory variables. Candidate cluster numbers from k = 2 to k = 6 were evaluated using within-cluster sum of squares, the Calinski-Harabasz index, an approximate centroid-based silhouette measure, minimum cluster size and clinical interpretability. Because k = 2 had the strongest separation metric, k = 3 was treated as an exploratory, clinically interpretable refinement rather than a uniquely optimal solution. Each primary solution used 20 k-means + + random initializations and retained the solution with the lowest within-cluster sum of squares. Initialization sensitivity was evaluated in 30 independent

batches of 20 initializations. Agreement with the reference k = 3 solution was quantified using the adjusted Rand index. Clusters were reordered by increasing MIMIC in-hospital mortality only after unsupervised fitting, with phenotype 1 as the lowest-risk reference phenotype.

Local patients were assigned to the nearest MIMIC-derived phenotype centroid in standardized Euclidean space after applying the frozen MIMIC imputation medians, means and standard deviations. The local data were not re-clustered and local outcomes were not used during assignment. This design evaluated transportability of the MIMIC-defined laboratory states rather than de novo phenotype discovery in the external cohort.

Outcomes and statistical analysis

The primary outcome was in-hospital mortality. Secondary outcomes included CRRT/RRT use and ICU length of stay. In MIMIC-IV, CRRT was defined at the ICU-stay level by at least one derived CRRT record with system_active = 1 or a non-empty crrt_mode; this rule identified 1,183 patients. RRT was defined by dialysis_present = 1 in the derived RRT table. Invasive mechanical ventilation was excluded from principal cross-cohort comparisons because the local source variable was insufficiently complete for reliable harmonization. Phenotype characteristics were summarized using medians and interquartile ranges or counts and percentages. Unadjusted mortality odds ratios were calculated using phenotype 1 as the reference. Multivariable logistic regression adjusted mortality associations for age, sex and SOFA score. Because SOFA overlaps with several organ dysfunction domains represented by the phenotype, the adjusted model was interpreted as a conservative assessment of residual association rather than a causal effect estimate.

To assess robustness of the k = 3 solution, 100 bootstrap resampling iterations were performed, each using 80% of the MIMIC-IV cohort. In each iteration, k = 3 clustering was repeated and clusters were matched to the original phenotype centroids. Stability was evaluated by whether phenotype 3 remained the highest-mortality and highest-CRRT cluster and whether monotonic gradients across phenotypes were preserved.

Results

Cohort characteristics

The MIMIC-IV screening rule identified 157 ICD-9/10 diagnosis codes and 43,092 hospital admissions with pneumonia-related diagnoses. After restricting to adult first ICU stays with an ICU length of stay of at least 24 hours, 16,933 ICU stays were included in the derivation cohort. In this cohort, 3,687 patients died during hospitalization, 2,025 received RRT and 1,183 received CRRT. The median ICU length of stay was 4.21 days. Re-execution of the archived cohort query reproduced all four screening counts and the mortality count exactly.

The external validation cohort included 412 patients with severe pneumonia admitted to the ICU of Longhua Hospital. The local cohort included 120 deaths and 65 patients who received CRRT. The local cohort was more clinically selected than the MIMIC-IV cohort because all patients met local severe pneumonia criteria (Fig. 1; Table 1).

Table 1
Cohort summary

Characteristic	MIMIC-IV derivation cohort	Chinese ICU validation cohort
Patients / ICU stays	16933	412
In-hospital deaths	3687	120
Mortality, %	21.8	29.1
RRT, n	2025	NA
CRRT, n	1183	65
Median ICU length of stay, days	4.21	15.0

Summary of the MIMIC-IV derivation cohort and Chinese ICU external assignment cohort. RRT was available in MIMIC-IV; the local analysis emphasized CRRT because this variable was consistently available. Local invasive ventilation was excluded because the source variable was insufficiently complete for reliable cross-cohort comparison.

Variable availability and model input

The core phenotype variables had low-to-moderate missingness in both cohorts. In MIMIC-IV, missingness was 0.6–0.7% for WBC, hemoglobin, platelet count, BUN, creatinine, sodium, potassium and chloride; 2.1% for calcium; 8.6% for PT; and 9.2% for APTT. In the local cohort, missingness was below 2% for all core variables except chloride, which was missing in 7.8%. In contrast, CRP, D-dimer, fibrinogen, NT-proBNP, CK-MB and troponin showed high early-window missingness in MIMIC-IV and were not included in the core model (Table 2).

Table 2
Missingness of core phenotype-defining variables

Variable	MIMIC missing %	Local missing %
WBC	0.7	0.2
Hemoglobin	0.6	0.0
Platelet	0.6	0.2
PT	8.6	1.2
APTT	9.2	1.7
BUN	0.6	0.2
Creatinine	0.6	0.2
Sodium	0.6	1.0
Potassium	0.6	1.2
Chloride	0.6	7.8
Calcium	2.1	1.2

Missingness of the 11 routine laboratory variables used for clustering and external phenotype assignment.

Selection of phenotype number

K-means clustering was evaluated from k = 2 to k = 6. The k = 2 solution had the highest approximate centroid-based silhouette and Calinski-Harabasz index. The prespecified exploratory k = 3 solution was retained because it separated a cytopenic/electrolyte-shift group from the relatively preserved group while maintaining adequate cluster sizes; it should not be interpreted as the uniquely optimal partition. The reference solution included 7,592 patients in phenotype 1, 6,282 in phenotype 2 and 3,059 in phenotype 3 after post hoc reordering by increasing mortality (Fig. 2A).

Laboratory characteristics of MIMIC-derived phenotypes

Phenotype 1 represented a relatively preserved hematologic–renal profile. Compared with the other phenotypes, it had higher hemoglobin and platelet count, lower PT and APTT, lower BUN and creatinine and relatively preserved calcium. This phenotype had the lowest severity and outcome risk.

Phenotype 2 showed a cytopenic and electrolyte-shift intermediate-risk profile. It was characterized by lower hemoglobin and platelet count than phenotype 1, modest coagulation abnormality, higher sodium and chloride, and lower calcium. Renal dysfunction was less prominent than in phenotype 3.

Phenotype 3 represented a renal dysfunction–coagulopathy high-risk profile. It showed markedly higher BUN and creatinine, prolonged PT and APTT, lower sodium, higher potassium, lower chloride, higher

WBC and low hemoglobin. This phenotype had the highest SOFA score and the highest CRRT rate (Fig. 3).

Outcomes across phenotypes in MIMIC-IV

In-hospital mortality increased across phenotypes. Mortality was 16.9% in phenotype 1, 22.6% in phenotype 2 and 32.2% in phenotype 3. CRRT use increased from 2.2% in phenotype 1 to 5.2% in phenotype 2 and 22.5% in phenotype 3. RRT use increased from 3.8% to 8.0% and 40.4%, respectively. The median SOFA score was 4 in phenotype 1, 6 in phenotype 2 and 8 in phenotype 3 (Table 3; Fig. 4A,B). Corresponding external-assignment outcomes are summarized in Table 4.

Table 3
MIMIC-IV phenotype outcomes

Phenotype	n	Mortality %	CRRT %	RRT %	Median ICU LOS	Median SOFA
1: relatively preserved hematologic–renal profile	7592	16.9	2.2	3.8	3.95	4
2: cytopenic/electrolyte-shift intermediate profile	6282	22.6	5.2	8.0	4.50	6
3: renal dysfunction–coagulopathy high-risk profile	3059	32.2	22.5	40.4	4.63	8

Outcomes and severity characteristics across MIMIC-derived phenotypes after reordering by increasing in-hospital mortality.

Table 4
External assignment phenotype outcomes

Phenotype	n	Mortality %	CRRT %	Median ICU LOS	Median SOFA	Median APACHE-II
1	18	16.7	0.0	11.0	4	12.5
2	358	27.7	11.7	15.0	6	19
3	36	50.0	63.9	11.5	8	21

Outcomes and severity characteristics after assigning local ICU patients to the fixed MIMIC-derived phenotype centroids.

Compared with phenotype 1, phenotype 2 had an unadjusted mortality OR of 1.443 (95% CI 1.326–1.570), while phenotype 3 had an unadjusted mortality OR of 2.342 (95% CI 2.126–2.580). After adjustment for age, sex and SOFA score, phenotype 2 was no longer independently associated with mortality (OR 0.972, 95% CI 0.887–1.064), whereas phenotype 3 retained a modest but significant association with mortality (OR 1.193, 95% CI 1.068–1.332; Table 5; Fig. 4C).

Table 5
Mortality odds ratios versus phenotype 1

Cohort	Phenotype	Model	OR	95% CI
MIMIC-IV	2 vs 1	Unadjusted	1.443	1.326–1.570
MIMIC-IV	3 vs 1	Unadjusted	2.342	2.126–2.580
MIMIC-IV	2 vs 1	Age + sex + SOFA adjusted	0.972	0.887–1.064
MIMIC-IV	3 vs 1	Age + sex + SOFA adjusted	1.193	1.068–1.332
Local	2 vs 1	Unadjusted	1.698	0.520–5.541
Local	3 vs 1	Unadjusted	4.429	1.175–16.693
Local	2 vs 1	Age + sex + SOFA adjusted	1.025	0.274–3.845
Local	3 vs 1	Age + sex + SOFA adjusted	1.855	0.419–8.214

Unadjusted and age-, sex- and SOFA-adjusted mortality odds ratios in the derivation and external validation cohorts.

Bootstrap stability analysis

In 100 bootstrap resampling iterations, phenotype 3 was identified as the highest-mortality cluster in 99% of iterations and as the highest-CRRT cluster in 99% of iterations. A strict monotonic CRRT gradient across phenotypes 1, 2 and 3 occurred in 85% of iterations, whereas a strict monotonic mortality gradient occurred in 45% of iterations. These findings indicate that the renal dysfunction–coagulopathy phenotype was stable as a high-risk group, while the boundary of the intermediate phenotype was less stable (Table 6; Fig. 2B).

Table 6
Bootstrap stability of the k = 3 solution

Stability metric	Result
Bootstrap iterations	100
Sample fraction per iteration	80%
Phenotype 3 highest mortality	99%
Phenotype 3 highest CRRT	99%
Strict monotonic mortality gradient P1 < P2<P3	45%
Strict monotonic CRRT gradient P1 < P2<P3	85%

Stability metrics from 100 bootstrap iterations using 80% MIMIC-IV resampling per iteration.

Initialization sensitivity analysis

Rebuilding the derivation cohort directly from MIMIC-IV reproduced 16,933 eligible ICU stays and 3,687 deaths. A separately implemented k-means procedure produced a closely concordant three-group solution (7,602, 6,269 and 3,062 patients after outcome ordering), with mortality of 16.9%, 22.6% and 32.2%, respectively. Across 30 independent analyses using 20 k-means + + initializations per analysis, every retained solution had an adjusted Rand index above 0.99 relative to the reference solution (minimum 0.993; median 0.993). In contrast, single-initialization runs occasionally converged to inferior local solutions, supporting the use of multistart fitting.

External validation in the Chinese ICU cohort

When the MIMIC-derived phenotype centroids were applied to the local ICU cohort, 18 patients were assigned to phenotype 1, 358 to phenotype 2 and 36 to phenotype 3. The low-risk phenotype was uncommon in the local cohort, consistent with the severe-pneumonia case definition used for external validation.

The local cohort showed a directionally consistent risk gradient. Mortality was 16.7% in phenotype 1, 27.7% in phenotype 2 and 50.0% in phenotype 3. CRRT use was 0.0%, 11.7% and 63.9%, respectively. Phenotype 3 also had the highest median SOFA score and APACHE-II score. Compared with phenotype 1, phenotype 3 had an unadjusted mortality OR of 4.429 (95% CI 1.175–16.693). After adjustment for age, sex and SOFA score, the estimate remained directionally elevated but was imprecise (OR 1.855, 95% CI 0.419–8.214), reflecting limited sample size in the local phenotype 1 group (Fig. 4C).

Discussion

In this MIMIC-IV derivation and Chinese external validation study, early routine laboratory tests identified an exploratory three-phenotype structure among ICU patients with pneumonia. The most robust and clinically important finding was not the complete ordering of all three groups, but a reproducible renal dysfunction–coagulopathy phenotype characterized by high BUN and creatinine, prolonged PT and APTT, electrolyte derangement and increased WBC. This phenotype showed the highest mortality, the highest SOFA score and a markedly increased need for CRRT in both MIMIC-IV and the external Chinese ICU cohort.

The study shifts the focus from conventional mortality prediction to clinically interpretable phenotyping. Many existing models in ICU pneumonia aim to predict death or organ support, but they do not necessarily explain how patients differ physiologically at ICU admission. Our findings suggest that routine laboratory tests can capture a high-risk patient state defined by renal dysfunction, coagulation abnormality and electrolyte disturbance. Fixed-centroid assignment identified a directionally similar high-risk group in the Chinese ICU cohort, although the small local phenotype 1 and phenotype 3 groups limit statistical confirmation.

The findings complement recent pneumonia subphenotyping studies. Multicohort hospital-acquired pneumonia work and prospective community-acquired pneumonia analyses have shown that pneumonia

contains reproducible clinical or biological subgroups with different outcomes [8, 9]. Our study addresses a different but practical question: whether a transportable phenotype can be derived from early routine ICU laboratory tests alone and externally assigned to a smaller real-world Chinese ICU cohort. This distinction is important because routine-laboratory phenotyping may be easier to reproduce than frameworks requiring research cytokines, omics measurements or inconsistently ordered biomarkers.

The renal dysfunction–coagulopathy phenotype may be clinically meaningful. Acute kidney injury, azotemia, electrolyte derangement and coagulation abnormalities are common manifestations of severe systemic illness and may reflect impaired perfusion, endothelial injury, inflammatory activation and multiorgan dysfunction. The markedly higher CRRT rate in phenotype 3 supports the clinical relevance of this laboratory pattern. Although phenotype 3 remained associated with mortality after adjustment for SOFA in MIMIC-IV, the effect size was attenuated, suggesting that part of its prognostic impact overlaps with acute organ dysfunction severity.

An important strength of this work is the use of routine laboratory variables rather than inconsistently measured biomarkers. CRP, procalcitonin, D-dimer, fibrinogen and natriuretic peptides may be informative in selected contexts, but they had high missingness in the early MIMIC-IV ICU window or were not consistently available. By focusing on widely measured variables and freezing all preprocessing parameters before local assignment, the phenotype framework may be easier to transport across hospitals and databases. This design does not, however, establish superiority over SOFA or simple laboratory thresholds.

The clinical implication is not that phenotype membership should immediately guide treatment, but that it may provide a reproducible way to describe high-risk physiologic states at ICU admission. In practice, a renal dysfunction–coagulopathy phenotype could help identify patients who warrant closer renal, coagulation and internal-environment monitoring, earlier multidisciplinary review and enrichment for future phenotype-stratified studies. These potential uses require prospective validation before clinical deployment.

The external assignment results also highlight important differences between cohorts. Phenotype 1 was uncommon in the local cohort, likely because the local dataset specifically enrolled patients meeting severe pneumonia criteria, whereas the MIMIC-IV cohort was based on ICU admission with pneumonia-related ICD diagnoses. The 18/358/36 local phenotype distribution produced imprecise adjusted estimates and precluded a strong claim of statistical replication. Nevertheless, phenotype 3 remained identifiable and showed the expected high-risk pattern. This finding supports directional transportability while indicating that phenotype prevalence, assignment distances and outcome associations may vary across healthcare systems and inclusion criteria.

This study has limitations. First, it is retrospective and subject to residual confounding, measurement error and center-specific treatment practices. Second, MIMIC-IV pneumonia was defined using ICD diagnosis descriptions, whereas the local cohort used clinical criteria for severe pneumonia; this

difference creates selection and case-mix heterogeneity. Third, k-means assumes approximately spherical clusters in standardized Euclidean space and remains sensitive to preprocessing and initialization. Multistart fitting reduced initialization instability, but alternative consensus clustering or latent-class analyses could yield different intermediate partitions. Fourth, k = 2 had stronger separation metrics than k = 3. The three-group solution was selected as an exploratory clinical refinement, and the intermediate phenotype should be interpreted descriptively rather than mechanistically. Fifth, the local cohort was single-center and small, with only 18 and 36 patients assigned to phenotypes 1 and 3. Its adjusted mortality estimates were consequently imprecise. Sixth, SOFA shares renal and coagulation domains with the laboratory phenotype; adjustment therefore attenuates the association and cannot establish independent biology or added clinical utility. Seventh, CRRT use may reflect both illness severity and center-specific practice. Eighth, additional multicentre validation, for example in eICU-CRD, would strengthen generalizability. Finally, phenotype membership should not be interpreted as a treatment recommendation without prospective validation.

Conclusions

In conclusion, early routine laboratory tests identified an exploratory renal dysfunction-coagulopathy high-risk phenotype among ICU patients with pneumonia. This phenotype was associated with higher mortality and substantially higher CRRT use in MIMIC-IV and showed directionally consistent fixed-centroid assignment in a Chinese ICU cohort. The findings support routine-laboratory phenotyping as a reproducible framework for describing ICU pneumonia heterogeneity and justify further multicentre and prospective validation before clinical deployment.

Abbreviations

APACHE-II
Acute Physiology and Chronic Health Evaluation II
APTT
activated partial thromboplastin time
BUN
blood urea nitrogen
CI
confidence interval
CK-MB
creatine kinase-MB
CRP
C-reactive protein
CRRT
continuous renal replacement therapy
ICU
intensive care unit

MIMIC-IV
Medical Information Mart for Intensive Care IV
NT-proBNP
N-terminal pro-B-type natriuretic peptide
OR
odds ratio
PT
prothrombin time
RRT
renal replacement therapy
SOFA
Sequential Organ Failure Assessment
WBC
white blood cell count.

Declarations

Ethics approval and consent to participate

MIMIC-IV is a publicly available de-identified database. Access was obtained after completion of the required training and data use agreement. The local retrospective cohort was approved by the Ethics Committee of Longhua Hospital, Shanghai University of Traditional Chinese Medicine (approval No. JXC72023030). Because this study used routinely collected retrospective clinical data, the requirement for informed consent was waived.

Consent for publication

Not applicable.

Use of generative AI and AI-assisted technologies

During manuscript revision, the authors used OpenAI Codex to assist with language editing, document formatting and preparation of reproducible analysis code. The authors critically reviewed and edited all AI-assisted outputs, verified the analyses against the source data and take full responsibility for the integrity and content of the manuscript. The AI tool is not listed as an author and was not used to make autonomous clinical or authorship decisions.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author Contribution

Y.L.Z. conceived the study question, coordinated the analysis workflow, interpreted the results and drafted the manuscript. S.H.Z. contributed to clinical data review, literature interpretation and revision of the Methods and Discussion. J.Y.L. contributed to data extraction, data cleaning, variable harmonization, statistical analysis and manuscript revision. X.L.W. contributed to data extraction, data cleaning, variable harmonization, statistical analysis and manuscript revision. Q.G. contributed to local database checking, outcome verification and table preparation. R.C. contributed to clinical interpretation, data quality review and critical revision of the manuscript. Y.Y.Z. provided overall supervision, secured study oversight, interpreted the findings and critically revised the manuscript. Y.L.Z., S.H.Z. and J.Y.L. contributed equally. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

MIMIC-IV is available through PhysioNet to credentialed users who complete the required training and data use agreement. The local ICU dataset generated and analyzed during the current study contains patient-level clinical information and is not publicly available because of privacy and institutional restrictions. De-identified analytic data and aggregate results may be made available from the corresponding author upon reasonable request and with appropriate institutional approval.

Code availability

The executable SQL and analysis scripts used to reconstruct the MIMIC-IV cohort, audit cohort counts, repeat k-means fitting and generate de-identified supporting tables are included in the reproducibility directory accompanying the submission package. Code for local-data harmonization and fixed-centroid assignment may be shared by the corresponding author after removal of institutional paths and patient identifiers, subject to institutional approval. Patient-level local data are not included.

Trial registration

Clinical trial number: not applicable.

References

1. Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023;10:1. 10.1038/s41597-022-01899-x.
2. Maslove DM, Tang B, Shankar-Hari M, Lawler PR, Angus DC, Baillie JK, et al. Redefining critical illness. Nat Med. 2022;28:1141–8. 10.1038/s41591-022-01843-x.
3. Mori M, Krumholz HM, Allore HG. Using latent class analysis to identify hidden clinical phenotypes. JAMA. 2020;324:700–1. 10.1001/jama.2020.2278.
4. Sinha P, Delucchi KL, McAuley DF, O'Kane CM, Matthay MA, Calfee CS. Development and validation of parsimonious algorithms to classify acute respiratory distress syndrome phenotypes: a secondary analysis of randomised controlled trials. Lancet Respiratory Med. 2020;8:247–57. 10.1016/S2213-2600(19)30369-8.
5. Maddali MV, Churpek M, Pham T, Rezoagli E, Zhuo H, Zhao W, et al. Validation and utility of ARDS subphenotypes identified by machine-learning models using clinical data: an observational, multicohort, retrospective analysis. Lancet Respiratory Med. 2022;10:367–77. 10.1016/S2213-2600(21)00461-6.
6. Sinha P, Kerchberger VE, Willmore A, Chambers J, Zhuo H, Abbott J, et al. Identifying molecular phenotypes in sepsis: an analysis of two prospective observational cohorts and secondary analysis of two randomised controlled trials. Lancet Respiratory Med. 2023;11:965–74. 10.1016/S2213-2600(23)00237-0.
7. Wendel Garcia PD, Caccioppola A, Coppola S, Pozzi T, Ciabattoni A, Cenci S, et al. Latent class analysis to predict intensive care outcomes in acute respiratory distress syndrome: a proposal of two pulmonary phenotypes. Crit Care. 2021;25:154. 10.1186/s13054-021-03578-6.
8. Martin FP, Poulain C, Haitsma Mulier J, Motos A, Gourain V, Ogan I, et al. Identification and validation of robust hospital-acquired pneumonia subphenotypes associated with all-cause mortality: a multi-cohort derivation and validation. Intensive Care Med. 2025;51:692–707. 10.1007/s00134-025-07884-3.
9. Coston TD, Poolchanuan P, Ross JE, Xia L, Zelnick LR, Hantrakun V, et al. Derivation and external validation of community-acquired pneumonia subphenotypes in Southeast Asia: a secondary analysis of prospective cohort studies. eClinicalMedicine. 2025;90:103572. 10.1016/j.eclinm.2025.103572.
10. Eickelberg G, Sanchez-Pinto LN, Kline AS, Luo Y. Transportability of bacterial infection prediction models for critically ill patients. J Am Med Inform Assoc. 2023;30:1874–85. 10.1093/jamia/ocad174.
11. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22:707–10. PMID:8844239.

12. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Critical Care Medicine., von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology statement: guidelines for reporting observational studies. PLoS Medicine. 2007;4:e296. 10.1371/journal.pmed.0040296
13. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. 10.1136/bmj-2023-078378.
14. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data. 2018;5:180178. 10.1038/sdata.2018.178.

Figures

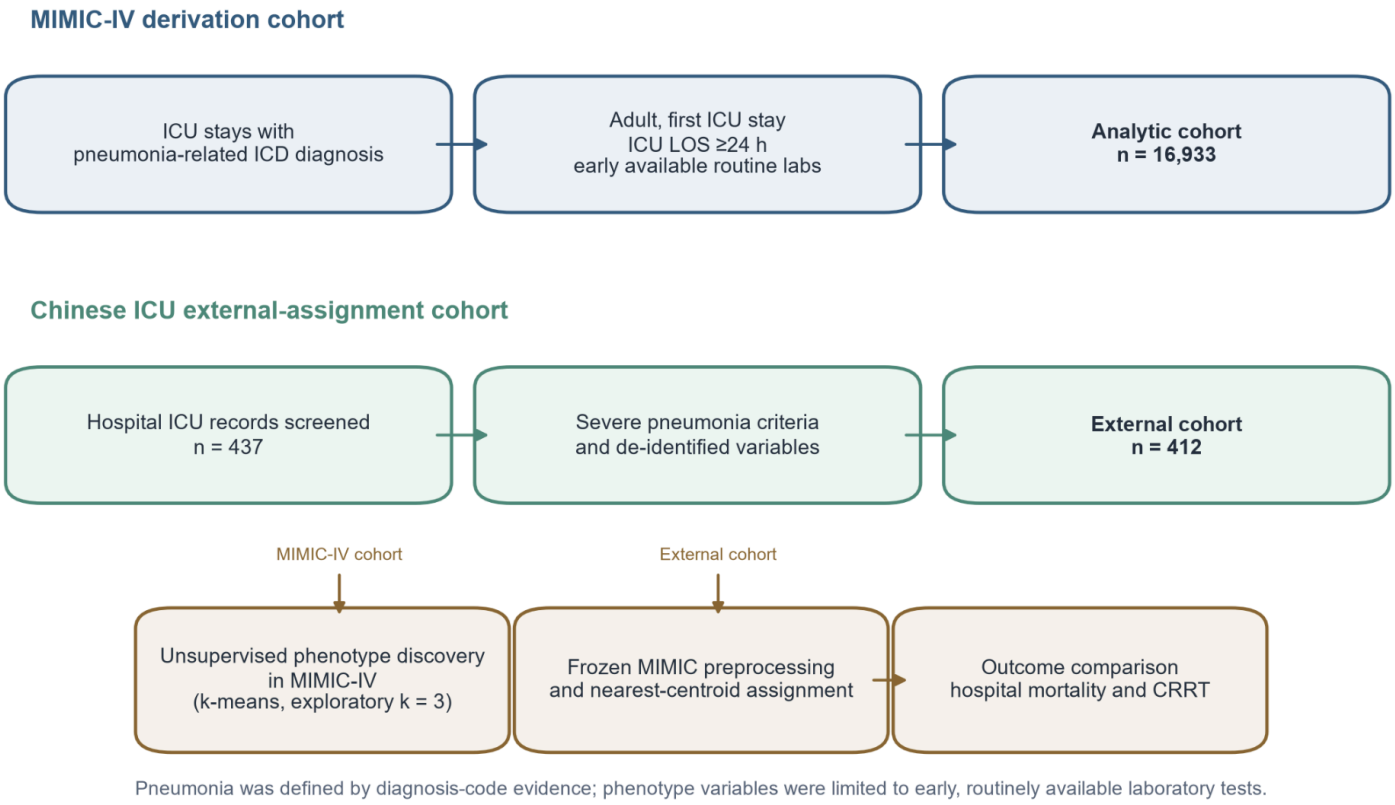


Figure 1

Study flow and fixed-centroid external-assignment design

Adult ICU patients with pneumonia were identified in MIMIC-IV and used for phenotype discovery. A de-identified Chinese ICU cohort was assigned to the frozen MIMIC-derived phenotype centroids without local re-clustering. Phenotype-defining variables were restricted to early routine laboratory tests.

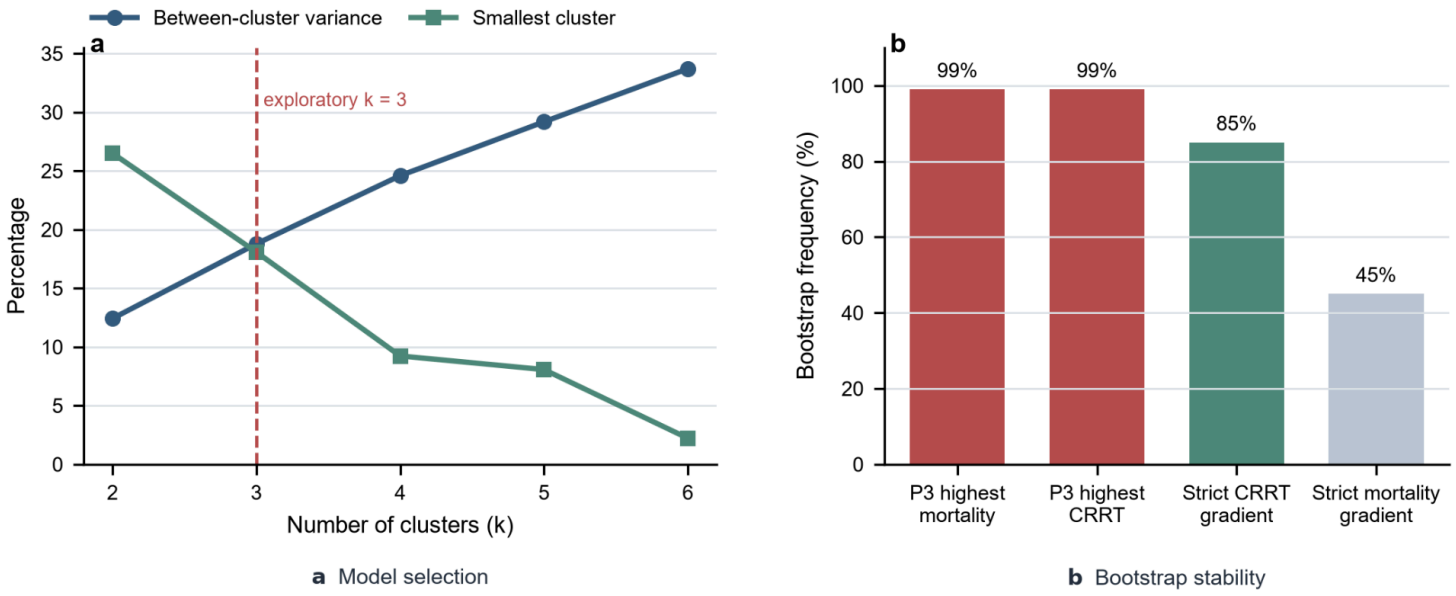


Figure 2

Phenotype number selection and bootstrap stability

a, K-means solutions from k=2 to k=6 were evaluated using between-cluster variance and minimum cluster size. Although k=2 had stronger separation metrics, k=3 was retained as an exploratory clinical refinement because it preserved adequate cluster size while providing clinically interpretable separation. b, Bootstrap stability analysis showed that phenotype 3 was the highest-mortality and highest-CRRT cluster in 99% of iterations, supporting the renal dysfunction-coagulopathy phenotype as the most reproducible finding.

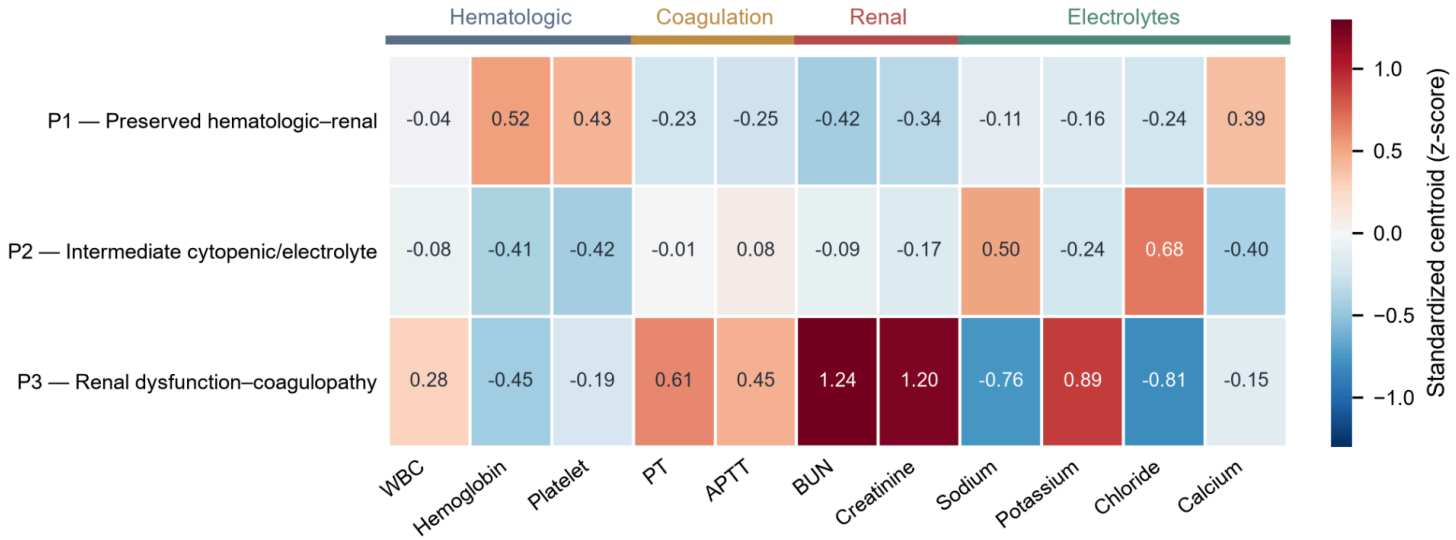


Figure 3

Standardized laboratory centroids of MIMIC-derived phenotypes

Heatmap values are z-score centroids derived from the MIMIC-IV cohort. Phenotype 3 was characterized by higher BUN and creatinine, prolonged PT and APTT, higher potassium, lower sodium and chloride, and lower hemoglobin, consistent with a renal dysfunction-coagulopathy high-risk phenotype.

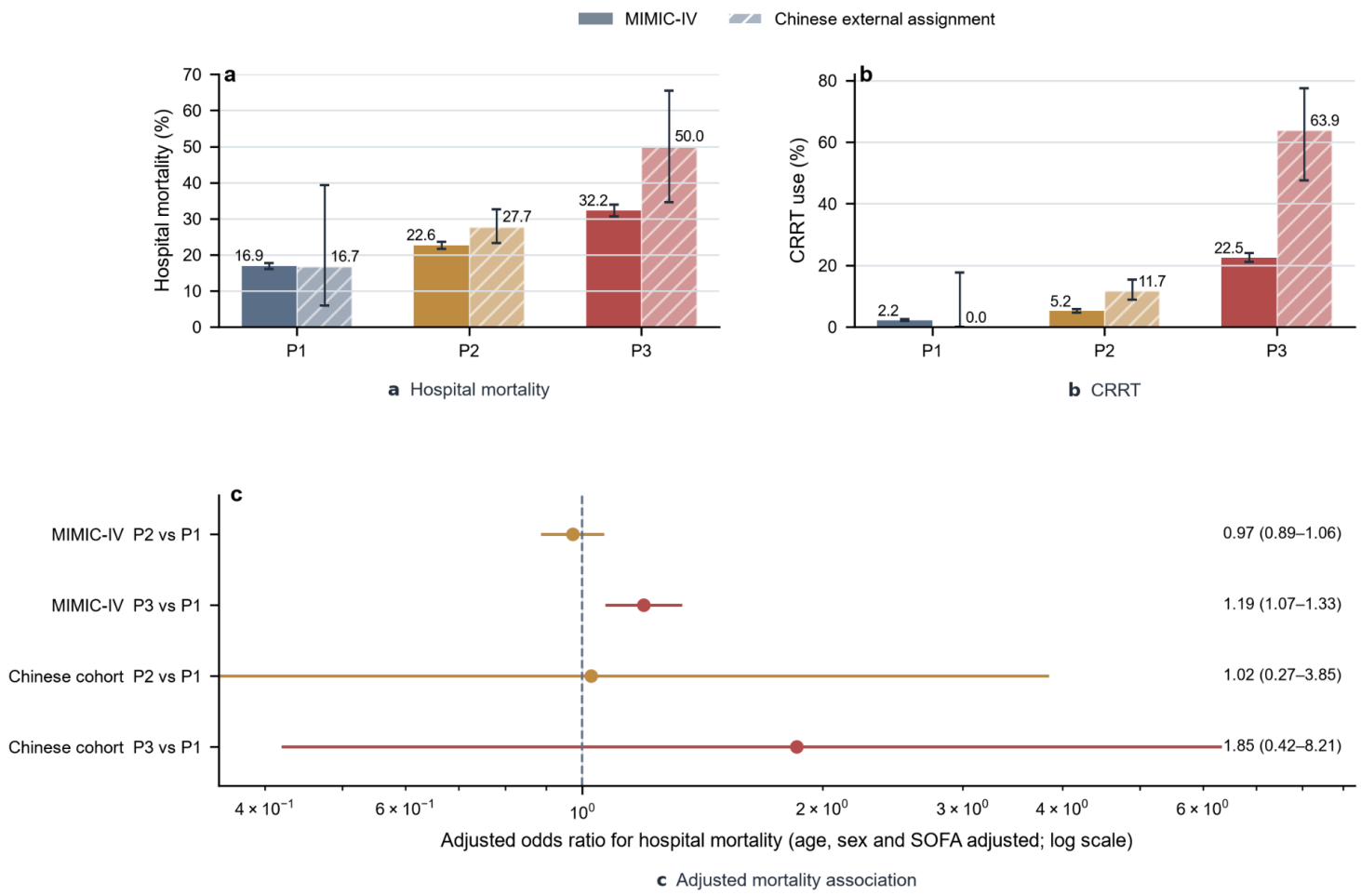


Figure 4

Outcomes across phenotypes

a, In-hospital mortality across phenotypes in MIMIC-IV and the Chinese fixed-centroid external-assignment cohort. b, CRRT use across phenotypes in both cohorts. Bars show percentages; error bars show Wilson 95% confidence intervals. Solid bars indicate MIMIC-IV and hatched bars indicate the Chinese cohort; bar colors distinguish phenotypes. c, Adjusted odds ratios for in-hospital mortality compared with phenotype 1. Models were adjusted for age, sex and SOFA score. Estimates are shown on a log scale with 95% confidence intervals.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterialv11.docx